AD-A036 579 WYLE LABS HUNTSVILLE ALA F/G 12/1

REGULATORY EFFECTIVENESS METHODOLOGY. PHASE I RESEARCH.(U)

JUL 76 S B COHEN DOT-CG-42333-A

UNCLASSIFIED MSR-76-42 USCG-D-7-77 NL

1 OF 2

AD A036579

ADA036579

49643

Report No. CG-D 7-77

REGULATORY EFFECTIVENESS METHODOLOGY

PHASE I RESEARCH

13

JULY 1976

FINAL REPORT

Document is available to the U.S. public through the National Technical Information Service, Springfield, Virginia 22161.

Prepared for

U. S. DEPARTMENT OF TRANSPORTATION
UNITED STATES COAST GUARD
Office of Research and Development
Washington, D.C. 20590

NOTICE

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
(18) US CG-D-7-77		

4. Title and Subtitle	5. Report Date
(6) REGULATORY EFFECTIVENESS METHODOLOGY. PHASE I RESEARCH	(11) July 1976
	6. Performing Organization Code (12) 96p.
7. Author(s)	8. Performing Organization Report No.
(10) Stephen B. Cohen	(14) MSR-76-42
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
Wyle Laboratories P.O. Box 1008 Huntsville, Alabama 35807	
	11. Contract or Grant No. (15) DOT-CG-42333-A T.O. 1
	13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address	(9) Final Report. July 1975-July 1976.
U.S. Department of Transportation United States Coast Guard Office of Research and Development Washington, D.C. 20590	14. Sponsoring Agency Code USCG (G-DSA)

15. Supplementary Notes

16. Abstract

Regulatory effectiveness methodologies are discussed in the context of recreational boating safety. Guidelines for developing data base models are given. Methods for predicting the expected benefits of a regulation are presented. Included are methods for handling unequally weighted probabilities (multi-state benefit analysis) and new methods of adjusting benefit computations to account for unreported accidents and victims. Forecasting problems are discussed. New techniques for assessing (tracking) post-regulation benefits are presented.

405 950

17. Key Words	18. Distribution Statement
Recreational Boating, Regulatory Effectiveness, Regulatory Benefits, Benefit Prediction, Benefit Assessment Safety.	Document is available to the U.S. public through the National Technical Information Service, Springfield, Virginia 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	93	

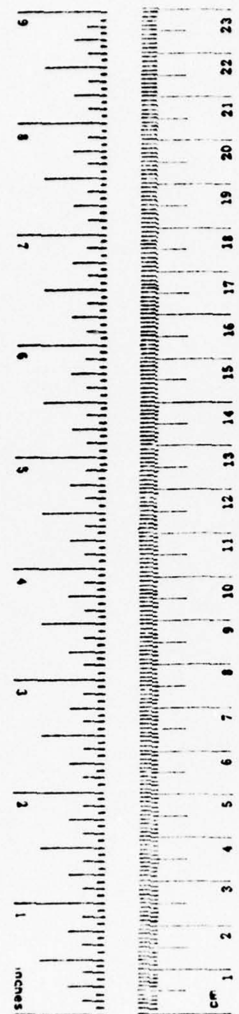
Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

METRIC CONVERSION FACTORS

Approximate Conversions to Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
in	inches	*2.5	centimeters	cm
ft	feet	30	centimeters	cm
yd	yards	0.9	meters	m
mi	miles	1.6	kilometers	km
		AREA		
in^2	square inches	6.5	square centimeters	cm^2
ft^2	square feet	0.09	square meters	m^2
yd^2	square yards	0.8	square meters	m^2
mi^2	square miles	2.6	square kilometers	km^2
	acres	0.4	hectares	ha
		MASS (weight)		
oz	ounces	28	grams	g
lb	pounds	0.45	kilograms	kg
	short tons (2000 lb)	0.9	tonnes	t
		VOLUME		
tsp	teaspoons	5	milliliters	ml
Tbsp	tablespoons	15	milliliters	ml
fl oz	fluid ounces	30	milliliters	ml
c	cups	0.24	liters	l
pt	pints	0.47	liters	l
qt	quarts	0.95	liters	l
gal	gallons	3.8	liters	l
ft^3	cubic feet	0.03	cubic meters	m^3
yd^3	cubic yards	0.76	cubic meters	m^3
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5/9 (after subtracting 32)	Celsius temperature	°C

*1 in = 2.54 (exactly).



Approximate Conversions from Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
mm	millimeters	0.04	inches	in
cm	centimeters	0.4	inches	in
m	meters	3.3	feet	ft
m	meters	1.1	yards	yd
km	kilometers	0.6	miles	mi
		AREA		
cm^2	square centimeters	0.16	square inches	in^2
m^2	square meters	1.2	square yards	yd^2
km^2	square kilometers	0.4	square miles	mi^2
ha	hectares (10,000 m^2)	2.5	acres	
		MASS (weight)		
g	grams	0.035	ounces	oz
kg	kilograms	2.2	pounds	lb
t	tonnes (1000 kg)	1.1	short tons	
		VOLUME		
ml	milliliters	0.03	fluid ounces	fl oz
l	liters	2.1	pints	pt
l	liters	1.06	quarts	qt
l	liters	0.26	gallons	gal
m^3	cubic meters	35	cubic feet	ft^3
m^3	cubic meters	1.3	cubic yards	yd^3
		TEMPERATURE (exact)		
°C	Celsius temperature	9/5 (then add 32)	Fahrenheit temperature	°F

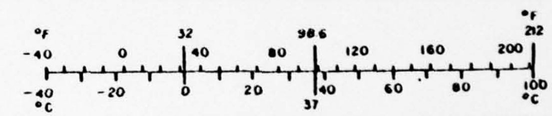


TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.0 INTRODUCTION

Increasing public and private interest in government regulation of various sectors of the economy and in areas of public activities has resulted in greater attention being paid to the effects of such regulation. Government agencies are now largely expected to justify proposed regulations and policies with studies of the anticipated benefits and costs, monetary and otherwise, of such proposals.

As a federal agency, the United States Coast Guard must likewise justify regulatory proposals. Regulation in the area of pleasure boating is a particularly sensitive issue. Not only is there likely to be public opposition to the regulation of recreational activities, but the economic impact on the boat manufacturing industry also must be carefully considered.

Because of this sensitivity, the Coast Guard must weigh its decisions carefully, perhaps more carefully than those of other agencies which have to deal with only a few large manufacturers. The Coast Guard requires, therefore, a relatively sophisticated methodology for assessing the effectiveness of its regulatory proposals.

Wyle Laboratories has been given the task of developing for the Coast Guard a general regulatory effectiveness methodology. (Regulation in this instance means not only mandatory requirements of boaters or manufacturers, but also education programs, changes in enforcement policies, etc. The word "regulation" will be used in this broad sense in this report.) This methodology must include techniques both for estimating the benefits and costs that would result from a contemplated regulation and for assessing or tracking the benefits and costs resulting from a regulation after it has been implemented.

Although costing methods can be quite complicated, they are relatively well-defined as compared with methods for benefit determination. This is due in large measure to differences in the types of data available. Manufacturers usually have detailed knowledge of their costs and can make good engineering judgments of what costs would be involved in the modification of an existing product or the manufacturing of a new one. Also, after a modified or new product is in the marketplace, there is good data on the actual costs incurred.

Probably the greatest difficulty in cost analysis is in projecting sales, material and labor costs, and factors related to the value of money, such as interest rates. These aspects are related to social changes and, thus, are related to similar problems in making benefit predictions.

The analysis of benefits, however, involves additional problems. Full data on the number of accidents, the number of victim recoveries and a host of more specific factors are not available. Some of this data could not be obtained even if a perfect accident reporting system were possible. For instance, there would be no way of determining the exact number of situations in which an accident would occur were it not for the effects of a regulation. At first, one might attempt to merely compare the number of accidents or fatalities occurring before the promulgation of a regulation with the number after the regulation went into effect. However, such an analysis does not take into account the unknown change in boating activity, including accidents and fatalities, that would have occurred if the regulation had not been adopted. While there is no means of absolutely determining exactly what such changes would have been, it should nevertheless be possible to develop techniques for evaluating regulatory benefits incorporating methods for estimating such changes.

As investigation of some of the methods heretofore used in regulatory effectiveness studies progressed, it became clear that it would be necessary to devise some new benefit determination methods which would be especially applicable to the sort of accident data available to the Coast Guard. Thus, a substantial part of Wyle's effort in this task was devoted to developing new methods for predicting and assessing (tracking) benefits. These methods are described in Sections 3.0 and 4.0 of this report.

In addition to having appropriate mathematical techniques for use in determining benefits, it is also necessary that available data be properly evaluated and structured so as to be amenable to mathematical and other analyses. This procedure is often referred to as modeling, although it actually is the development of a data base and not a true mathematical modeling effort. The Accident Recovery Model (Reference 1), developed during the same period that the work on this task was being performed, is a good example of such a data base. As a result of the experience

gained in developing ARM and other models, as well as work in this task, it has been possible to develop a set of guidelines for data base model development using Coast Guard accident data. These guidelines are presented in the next section.

It should be recognized that the methods presented in this report are preliminary in nature. Ongoing research will result in the refinement of these methods or the replacement of them by better ones. Furthermore, the reader should be aware that the methods are not necessarily presented in this report in the same order in which they might be applied. The reader is referred to Section 5.1 for a description of a possible procedure.

2.0 DATA EVALUATION AND STRUCTURING (MODELING)

2.1 Introduction

In order for the Coast Guard to determine how effective a contemplated regulation might be or a promulgated regulation has been, it is necessary for it to have information on how many accidents and victims the regulation might affect. In order to obtain this information, it is necessary that a structured analysis of accidents be performed. Such data analyses may be classified into two categories:

- analyses performed to determine the effectiveness of a specific regulation, and
- analyses performed on a general class of accidents or accident recovery for use in evaluating any regulation designed to affect the class.

An example of the first type of analysis is that performed by Operations Research Inc. in the analysis of bridge-to-bridge radio telephones (Reference 2). In that analysis, a structured series of questions called a Casualty Analyses Gauge (CAG) was used to evaluate accident reports to determine if the use of bridge-to-bridge radio telephones possibly could have prevented the accident, or if they were used and should have prevented the accident, but didn't. A sample of collision accident reports was chosen and each accident was analyzed using the CAG. The results were then summarized by two criteria: 1) radio telephone used or not used, and 2) radio telephone potentially could or could not have prevented the accident. Statistical methods were then applied to the data in an attempt to determine if the regulation requiring bridge-to-bridge radio telephones was beneficial in reducing accident occurrences.

The Accident Recovery Model, ARM (Reference 1), is an example of the second type of accident analysis. A sample of all types of recreational boating accidents was chosen and analyzed using a structured series of trees and coding definitions designed to evaluate the post-accident actions of accident victims and the conditions they encountered. The results of the analysis were then adjusted to reflect the statistics for all reported recreational boating accidents as found in CG-357. This type of analysis not only can be used in evaluating the effectiveness of regulations, but it can also be used to determine if a problem area exists which might benefit from regulation.

Before discussing any particular data analysis in depth, it seems appropriate to give some general guidelines for developing such an analysis so that the evaluation of accident reports and the structuring of the data will be efficiently performed. We begin the discussion of this topic in the following section.

2.2 Initial Steps in Developing a Data Base Model

The suggested procedure described in the following paragraphs is the result of experience gained in the development of a number of data base models. It is expected that further experience will lead to refinements and/or changes in this suggested guideline.

As a first step, at least one and preferably two analysts should read a variety of accident reports covering the subject to be analyzed. The reports read should include all types from simple BARs to MIO reports to in-depth reports, including some with supplementary materials such as newspaper reports. This will give the analysts a better feeling for the nature of the subject (e.g., collisions) and for the types of information available to him. Naturally, it is best if at least one of the analysts has had previous experience in analyzing and in investigating boating accidents.

After the analysts become familiar with the accident type(s) and the kinds of data available, they should identify the factors they believe to be important. For example, in the case of victim recovery, the condition of the boat, PFD wear, and swimming ability would be a few of the important factors. A structured data analysis form can then be constructed which includes these factors. The structured form may be a series of relevant questions, an event tree, or both. Experience seems to indicate that event trees are very useful at this stage. They help organize and direct the analysts' thoughts and help assure consistent evaluations (codings) of accident reports. If carefully developed, they also help in coding incomplete data in the many instances when fully detailed reports are unavailable.

Once the analysts are satisfied with the structured data analysis form they have developed, they should use the form to evaluate a number of accident reports. Two analysts should be involved in this step, and at least 30 reports of various types (BAR to in-depth) should be examined. At this stage, the coders will likely encounter many problems. Important factors

may be found to have been left out of the data analysis form, decisions on what actually occurred in the accident will have to be made, the problem of missing data will have to be attacked and fuzzy definitions and questions will have to be made more precise. The analysts should interact closely at this stage, making modifications in their accident analysis techniques as the evaluation of the sampled accident reports progresses.

It is almost a certainty that the analysts will encounter at least a few accidents which can be evaluated without too much difficulty, but for which the evaluation does not seem to adequately describe the circumstances of the accident — the evaluation doesn't "feel" right. This is most likely to occur when the circumstances involved are complex. If this happens in only one or two cases, there probably is nothing wrong, at least nothing serious, provided it isn't discovered that an important factor has been left out of the structured analysis form.

However, if this problem occurs often, the analysts should discuss it thoroughly and attempt to make adjustments to their evaluation methods to reduce the problem. Discussion with a third party can be valuable in this circumstance.

After the analysts have completed this preliminary accident sample evaluation, they should examine the results, including the changes that have taken place in the structured analysis form. If they feel that they now have developed their evaluation method sufficiently to evaluate a large sample, they are ready to proceed to the next step. If they are still unsatisfied with their evaluation method, or if their preliminary accident analysis resulted in major changes in their evaluation criteria, then they should probably go through a second preliminary accident sample evaluation, using revised methods.

Once satisfied with their accident report evaluation method, the analysts are then ready to adopt a format for their major accident evaluation effort. This should be in the form of a data base. That is, data should be coded in such a manner that any needed item of data is retrievable. Thus, unlike ARM_3 (Reference 1) in which, for example, it is only possible to retrieve information on the swimming abilities of victims who have no flotation aids, the format should allow retrieval of all information, in this instance, on the swimming ability of all victims, or at least all those for whom swimming ability could possibly have influenced recovery. Data for each accident or accident victim (whichever is appropriate) should be coded separately;

rather than coding summary data (totals, averages, etc.) in place of the individual data. In this way, the data can be later manipulated in whatever manner desired and the data base can be expanded by coding additional accidents and/or victims.

The implementation of the data base approach will make the use of a single event tree impossible, or virtually so. Such a tree would have to be very large and redundant, having many branches with identical nodes, or it would not provide the flexibility needed in order for all of the data to be coded, e.g., the example of swimming ability cited above. It may be possible, however, to use a series of small trees, each covering a single aspect of the accident or recovery. Whichever method is used, a series of small trees or a series of individual questions, the analysts should check to be sure that the data is coded in such a manner that each piece of data is individually retrievable. Definitions and questions should then be very carefully and completely expressed in written form. Because data results may be significantly different for serious accidents as opposed to non-serious ones, the data coded for each accident or victim should include whether or not the accident involved a fatality, serious injury, or significant property damage, as well as weather and water conditions.

In selection a sample of accident reports, the analysts may very likely wish to stratify the sample according to some parameters rather than to select a completely random sample. For example, if the data will be used to investigate a pre- vs. post-regulatory activity, then the analysts will want to select reports of accidents from particular years. They may also wish to stratify the selection by such factors as accident type, boat type, etc. Whatever stratification criteria are selected, the analysts must decide on the number or proportion of reports to be selected in each category. Within these constraints, the report selection should be random.

Two, or preferably three, individuals should be used to evaluate and code the sampled accident reports. At least one of the original analysts should remain involved, either as one of the evaluators and coders, or at least as a reference source when questions arise. When such questions do arise, the clarifications should be provided to all coders. At least two individuals should independently evaluate and code every accident. The coding should be done directly onto computer coding sheets to eliminate transcription errors. A "hard" copy should be made of every evaluated report, and a file should be kept of all such reports with a reference

numbering system so that the computer coded data can be matched to the report. This will facilitate the rechecking of analyses if such should be necessary and will allow for the collection of additional data on each accident, when needed.

After the accident reports have been evaluated and coded by at least two individuals working independently, the results should be compared. Disagreements should be worked out in consultations among the coders involved, and if any uncertainties remain, they should be discussed in a meeting of all coders, including the analyst who has acted as a reference source. This analyst should also have randomly checked accident evaluations to insure that the reports are being properly evaluated. Any disagreements which cannot be worked out should be coded "unknown."

Two final caveats with regard to coding: There may be a strong desire to use the coding "not applicable" in some circumstances. This desire should be strongly restrained. Any use of "not applicable" should be restricted to very specific circumstances. It is better to code all data, using "unknown" if necessary, so that in the future an analyst can decide for himself if any data is not applicable. The second caveat regards the coding of corrections and final decisions when disagreements occur. One and only one copy of the coding sheets should be prominently labeled as the "Corrected Copy." This should help reduce the likelihood of some corrections or changes not being included because of more than one copy being used for agreed-on corrections.

2.3 Weighting the Sample Data

In order to adjust the frequencies in a data base of sampled accident reports to reflect Coast Guard statistics for all reported accidents, it is necessary to use weighting factors. Also, even randomly chosen or carefully stratified samples will show some variation in relative frequencies when compared with the actual population from which the sample is drawn. This variation can be reduced by careful choice of weighting factors. It should be emphasized that weighting factors are used only to adjust the accident sample to reflect certain known, well-defined accident population statistics, such as numbers of fatalities. They are not used to adjust for ill-defined, or unknown, but estimated or desired accident population characteristics.

One first must decide on the parameters on which weighting should be based. This will depend on which are felt to be most important. Fatalities almost certainly will be included. Accident type, boat type and boat length are also likely to be considered important. Whichever parameters are chosen, it will be necessary to have matrices of frequencies for all combinations of parameter values, each parameter corresponding to one dimension of the matrix. One of these matrices will be of population frequencies, the other of sample frequencies. A matrix of weights can then be obtained by dividing each population frequency by the corresponding sample frequency. Mathematically,

$$W_{ij} = \frac{P_{ij}}{S_{ij}}, \quad S_{ij} \neq 0,$$

where W_{ij} is the weight in the i,j - position of the weight matrix W, and P_{ij} and S_{ij} are the frequencies in the i,j - positions of the population and sample frequency matrices, P and S, respectively.

In certain instances, the matrix S of sample frequencies will contain a zero entry while the corresponding entry in the matrix of population frequencies P is non-zero. This normally will occur only when the population frequency entry is small, so that the likelihood of a sampled accident being chosen with the given characteristics is small. In such a case any value for the weight could be entered in W as that weight would never be used. It is suggested that, as one check on the weighting program, a very large negative number be entered. Then if the weight is incorrectly used, such use will be obvious from the negative adjusted frequencies obtained.

If an entry of S is zero while the corresponding entry P_{ij} of P is non-zero, the above weighting procedure will not include the frequency P_{ij} in arriving at adjusted frequency totals, and such totals will therefore be smaller than the actual frequency totals. If this is undesirable, the following alternatives are suggested:

- If the category with no sample data is considered important, the Coast Guard data bank can be searched for the relevant accidents in order to obtain the case numbers of the accidents. The accident reports can then be located and added to the sample. This will result in a revised matrix S of sample frequencies and a corresponding revised weight matrix W.

- To merely make a final adjustment of frequencies to bring the weighted sample frequencies up to the population frequencies, each computed weighted frequency can be multiplied by:

$$\alpha = \frac{\sum_{ij} P_{ij}}{\sum_{\substack{ij \\ S_{ij} \neq 0}} W_{ij} S_{ij}} .$$

  That is, by the ratio of the total population frequency to the adjusted sample frequency (adjusted by using weights derived from non-zero values of S_{ij}). In effect, this treats subcategories for which there are no sample reports as being an "average" of all sampled accidents.

- A procedure similar to the preceeding can be performed with the adjustments being made within only some categories (e.g., fires) rather than overall. In effect, this treats a subcategory (e.g., fires on sailboats) for which there are no sample reports as being an "average" of all sampled accidents in the category (fires) containing that particular subcategory.

The latter two procedures may both introduce considerable error if the accident subcategories involved have relatively large frequencies (P_{ij}) in the accident population or differ significantly from "average" sampled accidents.

The actual computation of the weights could be performed on the computer. If the Coast Guard data base of all reported and coded accidents is available in memory, the matrix P of accident population frequencies can be constructed by making tabulations of the number of population accidents corresponding to particular parameter values. Otherwise, the matrix P must be obtained from another source, probably CG-357, and entered into computer memory. A computer program is then used to tabulate the number of sample accidents corresponding to particular parameter values, thus constructing the matrix S, and finally the matrix W of weights is computed.

In the case of models in which victims are individually accounted for, the weighting will normally apply to victims rather than to accidents. No tabulation is kept on the total number of victims involved in reported accidents; only injured victims and fatalities are tabulated. The population frequencies must therefore be estimated from the sample victim frequencies. An overall estimate or a category-by-category estimate can be obtained using the numbers of people on board. Thus,

$$\text{estimated population victim frequency} = \left(\frac{\text{sample victim frequency}}{\text{sample boat frequency}}\right) \cdot \left(\begin{array}{c}\text{population}\\ \text{boat frequency}\end{array}\right).$$

The number of survivors or non-injured victims can then be estimated by subtracting from this number the number of fatalities or injured victims, respectively, adjusting where necessary for victims in the water who were not on a boat.

The computer program should contain a subroutine to apply the weights as part of a sorting routine in the following manner. In sorting, when an accident (or victim) meets the sort criteria, the appropriate weight is selected from W and the tabulation totals are incremented by this weight, rather than by "one." Thus, in effect, each sample accident (or victim) in the i,j - category is counted as W_{ij} population accidents (victims) in that category.

Because it is important to be able to update the data base and to account for changes in the accident population, the above described method of applying the weights is appropriate. The alternative would be to replace each accident (victim) entry in the data base with its appro-

priate weight and thus be able to tabulate weights directly. However, this would make data updating more difficult as well as requiring more storage space for the extra digits needed in each data entry. For the most efficient updating of data, a computer routine for recomputing the matrices P, S and W should be available.

2.4 Model Integration — Developing an Overall Data Base Model

One of the requirements of a complete program of regulatory effectiveness analysis is a data base model covering all important accident types as well as post-accident victim recovery. Data analyses of the type previously discussed should be performed in at least each of the following areas:

- accident recovery (the Accident Recovery Model currently under development),
- collisions, including groundings, and two-boat, fixed object and floating object collisions,
- capsizings, swampings, and sinkings,
- fires and explosions,
- falls overboard and within boat, and
- struck by boat or propeller.

Although separate data base models may be developed in each of these areas, they should be developed so as to be compatible, so that they can be integrated into a single data base covering all major aspects of boat accident and victim recovery. The requirements for compatibility and completeness include at least the following:

- coding all basic sample data including boat type, length, horsepower, etc.
- using standardized coding in all work so that, for instance, the same code is used for houseboats in all models
- using the same sample for coding recovery as is used for coding accident cause, and

- developing computer software (programs) with sufficient generality so as to be applicable to all areas; or, as an intermediate requirement, thoroughly documenting individual programs so a program can be eventually developed which includes all features of the individual programs.

The advantage of having a unified data base is that information across accident types can be generated. For instance, a special model could be constructed covering all areas of accident cause and victim recovery in outboard boats. If, say, a powering regulation for outboards is contemplated, one would be able to examine all accident causes which might be related to powering.

Once model integration has been accomplished, there should be an effort to extend its usefulness in pre-regulation benefit prediction and post-regulation benefit assessment (tracking). Implementation of methods in these areas require the development of a supplementary data base as well as additional, supplementary programs. Descriptions of these are deferred to Sections 3.0 and 4.0.

3.0 BENEFIT PREDICTION

3.1 Introduction

There should be some indication that the promulgation of a contemplated regulation will result in the saving of lives, a reduction in injuries and/or a reduction in property damage. This section describes some techniques useful in predicting such benefits. It also includes a discussion of some of the problems involved in attempting such prediction. One certainty has emerged during research into benefit prediction and assessment techniques: unforeseen difficulties can appear in the most unexpected places when carefully analyzing an actual problem. Although these difficulties sometimes seem to occur almost by chance, they nevertheless result in a fuller understanding of the processes involved. For this reason, it is imperative that research in the areas of benefit prediction and assessment techniques not be performed in a vacuum, but rather that such research involve continuing application to real data, both to help validate the techniques developed and to broaden understanding of the complexities involved. When real data is not currently available, carefully developed synthetic data should be used to test the techniques.

Early in the research into regulatory effectiveness methodology, it was decided that benefit prediction could usually be separated into the following three phases:

- estimation of the effect that a regulation will have on the specific accident or recovery causative factor(s) it is designed to influence, as well as its effect, positive or negative, on other factors,
- estimation of the benefit which would accrue if the full effect on the influenced factor(s) were available immediately, and
- projection of the future benefit accruing from the regulation taking into account its implementation rate and projected boating trends.

The first phase would appear to require a separate study for each regulation or set of regulations under consideration. For example, to estimate the effect of an education program designed to increase PFD wear, it might be necessary to run tests using the educational material to determine its influence on boaters' behavior. Such experimentation along with other data such as that from ARM, would result in an estimation of the proportion of people who would be influenced to wear PFDs.

Once the effect of a regulation on the involved factor(s) has been estimated, it should be possible to use the data base model to estimate what the resultant benefit would be if the full effect of the regulation could be felt immediately. For instance, in the case of education to increase PFD wear, ARM would provide an estimate of the number of lives saved were it possible to immediately increase PFD wear by the amount determined in the first phase.

The third phase of benefit prediction involves forecasting what the future benefit will be. In effect, this involves these steps:

- generating a curve which predicts the change in the affected factor(s) over time,
- forecasting what changes in boating accident patterns would occur were the regulation not promulgated, and
- combining the foregoing with the benefit figure(s) derived in phase two.

In the following pages, each of these three phases of benefit prediction is discussed in more detail.

3.2 Estimation of the Effect of a Regulation on Accident or Recovery Causative Factors

Evidence of the possible need for a regulation is usually first discovered by the Coast Guard. Sources of evidence include analyses of data obtained from accident reports compiled by the Coast Guard from its own investigations and from investigations performed by state and local authorities. Other sources include the American Boat and Yacht Council (ABYC), other boater and consumer organizations, and the boating industry, including manufacturer defect reports.

Once the Coast Guard Office of Boating Safety has identified the possibility of a safety problem, it requests the Coast Guard Office of Research and Development to further investigate the situation. The first step is a cause identification phase. Statistical analyses of data base models such as described in Section 2.0 may be quite useful in both this stage and in the original problem identification stage. At least, such analyses should help narrow the area of investigation. However, research methods tailored to the particular problem will also be

performed. These will likely include further analysis of accident reports, in-depth accident investigations, and possibly surveys and experiments involving boaters and/or boater equipment designs.

The research results in the development of alternative safety enhancement program concepts. These program concepts are studied by the Office of Boating Safety which determines those for which advanced development research is desirable. Advanced development involves further research into the precise form of a program and into its expected benefits and costs. It is in this phase and the previous cause identification phase that estimates are made of the effect that a regulation or program will have an accident or recovery causative facators.

These estimates will be made by using expert judgment in conjunction with statistical data analyses, equipment performance tests, equipment design investigations, experiments, surveys, and prototype programs. For example, the effects that a possible modification of PFD approval standards would have on PFD wear could be estimated through analyses of studies of current PFD wear and the attitudes of boaters to probable new PFD designs. The estimated effects (e.g., a 20% wear rate increase) can then be used in making benefit estimates as described in the following sections.

3.3 "Full Effect" Benefit Estimation

In Section 3.2 we discussed the first phase of benefit prediction, concerning the determination of what effects a contemplated regulation would have on accident or recovery causative factors. In this section we discuss the problem of converting these effects (e.g., increased PFD wear) into initial benefit estimates.

At first glance this might seem to be a straight-forward, relatively simple process. Using the new factor estimates, say increase in PFD wear, one uses the data base model to calculate the change in accident frequencies or victim fatalities. Regrettably, the process is not so simple. Both mathematical problems and problems with insufficient data exist. Initial research indicated that these problems were sufficiently complex to deserve a significant portion of the task effort. Because of the complexities involved, a semi-chronological approach will be

taken in discussing the research, so that the reader will have an opportunity to see how problems and complexities developed. Furthermore, in order to illustrate the methods and problems with reasonably simple examples, synthetic (simulated) data will often be used. Although the methods presented were developed in conjunction with, and have been applied to, real data (see Reference 1) the use of real data in examples would often unnecessarily complicate them and/or obscure the principles being discussed.

3.3.1 Benefit Estimation Based Only on Reported Victims

In its most simplistic form, the calculation of benefits involves merely the mathematical transfer of a certain number of reported accidents or victims * from a "less desirable" state to a "more desirable" state. In the case of accident recovery, the Accident Recovery Model includes data on estimated annual numbers of reported victims and survivors who are in different states; that is, who have different survival factors associated with them. To describe the method, it is helpful to symbolize the quantities included. Let S_0 be a "less desirable" state, for example non-use of a PFD by a victim in the water. Let S_1 be a "more desirable" state, for example use of a PFD by a victim in the water. Suppose these states are disjoint; that is, no victim can be in both of them. Finally, let p_0 and p_1 be the probabilities of survival of victims in states S_0 and S_1, respectively. These probabilities may be calculated as:

$$p_0 = \frac{a}{b} \text{ and } p_1 = \frac{c}{d}, \qquad (3\text{-}1)$$

where a and c are the average annual number of reported accident survivors in states S_0 and S_1, and b and d are the total average annual number of reported victims in states S_0 and S_1, respectively. Note that $a \leq b$ and $c \leq d$.

---

* A victim is defined as any person involved in an accident, whether or not the person survives.

It is helpful to diagram those quantities as follows:

S_0	S_1
$\frac{a}{b} = p_0$	$\frac{c}{d} = p_1$.

To say S_0 is less desirable than S_1 means $p_0 < p_1$. Thus, for example, a regulation designed to increase PFD use, that is, to cause more victims to be in state S_1 and fewer to be in state S_0, should result in more victims surviving. The most simplistic approach to benefit calculation involves mathematically "transferring" victims from state S_0 to state S_1. Suppose research of the type described in Section 3.2 indicates that a fraction, r, of victims in state S_0 can be transferred to S_1. For example, perhaps an educational program could result in 30% (r = 0.30) of the population who would not otherwise use PFDs in an accident situation to use PFDs. The transferred victims would have a higher probability of recovery, p_1, than they otherwise would have. The benefit resulting from the regulation, once it became effective at the predicted level, could then be calculated. The benefit B_0 in average number of lives saved yearly can be calculated from:

- $$B_0 = rb\,(p_1 - p_0)\ . \qquad (3\text{-}2)$$

This formula is based on the assumptions:

- *All victims in state S_0 have an equal probability of being transferred to state S_1.*
- The recovery probability of a victim is determined solely by the state he is in.

The derivation of this formula may be found in Appendix A.

As an example, suppose the average annual number of reported victims in the water wearing PFDs is 2313 and of these victims 2267 survive, yielding 98% probability of recovery, while 4136 victims in the water do not use PFDs and of these 3598 survive, making their probability of recovery 87%. This may be diagramed as:

S_0	S_1
$\frac{a}{b} = \frac{3598}{4136} = 0.87 = p_0$	$\frac{c}{d} = \frac{2267}{2313} = 0.98 = p_1$

If research of the type described in Section 3.1 indicates that a contemplated regulation would eventually result in PFD wear for 30% of the victims who otherwise would not wear them, then the eventual benefit could be calculated as:

$$B_0 = 0.30\ (4136)\ (0.98 - 0.87) = 136 \text{ additional lives saved per year.}$$

Note that we assume there is no change in accident or victim recovery patterns except for the transfer effect of the regulation.

This method can also be used to predict the number of lives saved as the result of a standard designed to reduce the number of accidents by reducing the occurrence of a causative factor. Suppose that 151 collisions involving 1252 victims of whom nine die occur mainly as the result of steering malfunctions. *If it is estimated that a new standard will* halve the number of steering malfunctions, then the benefit in lives saved once the standard became fully effective would be:

$$\begin{aligned} B_0 &= 0.5\ (1252)\ (1.00 - 0.9928) \\ &= 4.5 \end{aligned}$$

where 1.00 is the probability of survival when there is no accident and $\frac{1252-9}{1252} = 0.9928$ is the survival probability of a collision victim in an accident involving a steering malfunction. Of course, this result could be obtained more simply as $B_0 = \frac{1}{2}(9) = 4.5$.

3.3.2 Benefit Estimation Taking Into Account Unreported Victims

A problem with the above method of benefit prediction almost immediately presented itself. The Coast Guard has no record of most non-fatal accidents and accident victims because most such accidents are never reported. Thus, there are many accidents and victims not accounted for in the statistics or the data base models, and the survival probabilities are higher than indicated because these victims are survivors. (Virtually all fatalities are reported.)

Two means of dealing with this problem were considered. The total number of victims, reported and unreported, in each state could be estimated and these numbers could be used to calculate survival probabilities and benefits. The other means considered involved developing error bounds on the calculated benefit. It was felt that this later approach would be more useful and would require less data estimation or the need for additional data collection through surveys, etc.

In order to develop error bounds on the calculated benefit B_0, it was first necessary to express the actual benefit B as a function of additional variables, defined as follows:

- Let x be the total average annual number of unreported accident victims in states S_0 and S_1, with α and β the fractions of these victims in state S_0 and in state S_1, respectively.

- As an immediate consequence of this definition we have that αx is the number of unreported victims in state S_0, βx is the number of unreported victims in state S_1 and $\alpha + \beta = 1$.

In order to make use of these variables, we make the assumption:

- All unreported accident victims are survivors.

This assumption is consistent with current estimates of the reporting rates for fatalities which indicate that virtually all fatalities are reported. With this assumption, the total numbers of victims (reported and unreported) in states S_0 and S_1 can be expressed as $b + \alpha x$ and $d + \beta x$, respectively, while the corresponding numbers of survivors are given by $a + \alpha x$ and $c + \beta x$.

Using these quantities to express the survival probabilities p'_0 and p'_1 for all victims in states S_0 and S_1, we have:

- $$p'_0 = \frac{a + \alpha x}{b + \alpha x} \text{ and } p'_1 = \frac{c + \beta x}{d + \beta x}. \qquad (3\text{-}3)$$

  This may be diagramed as:

S_0	S_1
$\frac{a + \alpha x}{b + \alpha x} = p'_0$	$\frac{c + \beta x}{d + \beta x} = p'_1$

The full effect benefit B expressed as the average annual number of lives saved is given by:

- $$B = B(x, \alpha) = r(b + \alpha x)(p'_1 - p'_0) \qquad (3\text{-}4)$$

where, as before, r is the fraction of victims "transferred" by the regulation from state S_0 to state S_1.

3.3.2.1 Testing for Positive and Minimum Benefits — One is interested, of course, in regulations which will cause boats or accident victims to be in more desirable states. That is, one desires positive benefits (lives saved). Mathematically, this means that $p'_0 < p'_1$, since we are considering S_0 as the state transferred from and S_1 as the state transferred to. Thus, we have as a condition for positive benefits:

- $$p'_0 < p'_1 \,^{*} \qquad (3\text{-}5)$$

or equivalently,

- $$(a + \alpha x)(d + \beta x) < (b + \alpha x)(c + \beta x), \ \beta = 1 - \alpha \qquad (3\text{-}6)$$

Formulas 3-3 and 3-4 can be used to calculate benefits directly if the quantities x and α can be estimated. If only range estimates of α or x can be made, then the error bounds discussed later can be used.

---

* Since the actual requirement for positive benefits is $p'_0 < p'_1$, it is theoretically possible to have $p'_0 < p'_1$ while concurrently having $p_0 > p_1$. However, this circumstance is highly unlikely and, in view of the dearth of data on unreported accidents and victims, the consideration of a regulation affecting conditions in which $p'_0 < p'_1$ but $p_0 > p_1$ is even more unlikely. Clearly, any such regulation would require the most careful study.

Before any significant amount of work is performed to attain estimates of α or x, however, some attempt should be made to determine if inequalities 3-5 are satisfied. Checking the inequality $p_0 < p_1$ (or equivalently, $ad < bc$) is easy provided the data base model contains the appropriate data. To check the inequality $p'_0 < p'_1$ at least rough variable estimates are needed. Table 1 contains alternate forms of inequalities 3-5 and 3-6 which may be easier to check. Note that as the two inequalities under condition A in the table are equivalent, if one is true then both are true, and if one is false then both are false, so it is only necessary to check one of them.

As the inequalities under condition A do not involve the variable x, it should be a relatively straightforward process to determine if this condition is met. Conditions B and C are somewhat more difficult to check as the variable x is also present.

To ease the problem of having to estimate both x and α, two methods are suggested. The first involves estimating one of these variables and using the estimate in computing the right side of inequality 3-9, 3-10 or 3-11, whichever is appropriate. The value obtained will be an upper or lower bound on the value of the remaining variable, if positive benefits are to result. One then must only decide if the remaining variable has a value below or above this bound, and, thus, if the inequality is satisfied and condition B or C is met. The same procedure can be used in testing condition A by rewriting inequality 3-8 as:

$$\alpha < \frac{b - a}{b + d - a - c} \quad . \tag{3-12}$$

The second method involves graphically testing inequality 3-11 by graphing the curve:

$$\alpha = \frac{1}{(b - a) + (d - c)} \left((bc - ad) \frac{1}{x} + (b - a) \right) , \tag{3-13}$$

(where $\alpha + \beta = 1$)

with α as the ordinate variable and x as the abscissa variable. One can then examine the graph to determine if any reasonable combination of x and α would fall above the curve implying inequality 3-11 might be false and, thus, that the regulation might result in negative benefits.

TABLE 1. TESTING FOR POSITIVE BENEFITS

If condition A, condition B, or condition C is met, then $p'_0 < p'_1$ and positive benefits will result from a regulation transferring accident victims from state S_0 to state S_1.

A. Both $p_0 < p_1$ and either of the following inequalities is true:

1. $\frac{d - c}{b - a} \leq \frac{\beta}{\alpha}$, or equivalently

$$\frac{\text{fatalities in state } S_1}{\text{fatalities in state } S_0} \leq \frac{\text{unreported victims in state } S_1}{\text{unreported victims in state } S_0}, \; 0 < \alpha < 1 \qquad (3\text{-}7)$$

2. $(\alpha)(b + d - a - c) \leq (b - a)$, or equivalently

$$\text{percent of unreported victims in state } S_0 \leq \text{percent of fatalities in state } S_0 . \qquad (3\text{-}8)$$

B. One of the following is true:

1. The two equivalent inequalities 3-7 and 3-8 are true and

$$x > \frac{ad - bc}{(b - a)\beta - (d - c)\alpha}, \; (b - a)\beta - (d - c)\alpha \neq 0 \qquad (3\text{-}9)$$

2. The two equivalent inequalities 3-7 and 3-8 are false and the following inequality is true:

$$x < \frac{bc - ad}{(d - c)\alpha - (b - a)\beta} . \qquad (3\text{-}10)$$

C. The following inequality is true:

$$\alpha < \frac{1}{b + d - a - c} \left((bc - ad) \frac{1}{x} + b - a \right) \qquad (3\text{-}11)$$

As an illustration, let us return to the earlier example on PFD wear. Recall that S_0 was taken as the state of non-PFD wear and S_1 as the state of PFD wear. Representing the quantities schematically, we have:

S_0	S_1
$\frac{a}{b} = \frac{3598}{4136} = 0.87 = p_0$	$\frac{c}{d} = \frac{2267}{2313} = 0.98 = p_1$.

Note that the inequality $p_0 < p_1$ is satisfied.

The first step required in order to take into account unreported accident victims (all assumed to be survivors) is to check that transferring both reported and unreported victims from S_0 to S_1 will produce positive benefits; that is, to check that $p'_0 < p'_1$. To do this, we first check condition A. Suppose a rough estimate of the fraction of unreported victims who wear PFDs is $\beta = 0.15$. Then $\alpha = 0.85$, and we can test inequality (3-7):

$$\frac{d - c}{\beta} = \frac{2313 - 2267}{0.15} = 306.7$$

$$\frac{b - a}{\alpha} = \frac{4136 - 3598}{0.85} = 632.9 \quad .$$

We find that inequality 3-7 is satisfied and, thus, condition A is met, meaning that $p'_0 < p'_1$.

For a second illustration, suppose that instead of the estimate of α as 0.85, we estimate that $\alpha = 0.95$. Testing condition A with this value, we might use inequality 3-8:

$$(\alpha)(b + d - a - c) = (0.95)(4136 + 2313 - 3598 - 2267) = 554.8$$

$$b - a = 4136 - 3598 = 538 \; .$$

Inequality 3-8 is not satisfied, so condition A is not met and we must proceed to check either condition B or condition C. Let us check condition B. Suppose we are unsure as to what value we should take for x . We can first calculate the value of the right side of inequality 3-10, noting that $\beta = 1 - \alpha = 0.05$:

$$\frac{bc - ad}{(d - c)\alpha - (b - a)\beta} = \frac{(4136)(2267) - (3598)(2313)}{(2313 - 2267)(0.95) - (4136 - 3598)(0.05)} = 62{,}746 \; .$$

A decision must now be made: Is inequality 3-10 true? That is, is x less than or greater than 62,746? If $x < 62,746$, then $p'_0 < p'_1$ and positive benefits (lives saved) will result from a regulation resulting in increased PFD wear. If $x \geq 62,746$, then $p'_0 \geq p'_1$ and no benefit or a negative benefit (lives lost) would result from such a regulation.

Should it be determined that $p'_0 \geq p'_1$ then any regulation having as its only effect the transfer of victims from state S_0 to S_1 should be counterproductive*. If it is found that $p'_0 < p'_1$, then such a regulation might be beneficial and further study is warranted. Even if it is determined that $p'_0 < p'_1$, it may be that the benefit derivable from a regulation is too small to justify the regulation's promulgation. Certainly, it is preferable to obtain a rough estimate of what the benefit may be before expending a large sum in further research. The benefit may be calculated from:

o $$B = r(b + \alpha x)(p'_1 - p'_0),$$

or the equivalent formula:

o $$B = \frac{r}{d + \beta x}\left(bc - ad + [(b - a)\beta - (d - c)\alpha]\, x \right). \qquad (3\text{-}14)$$

Of course, the same problem of estimating x and α remains. (Recall that $\beta = 1 - \alpha$, so an estimate of α yields an estimate of β.) Again, it is possible to ease this estimation problem. One means of doing so is to estimate one of the variables, α or x, and decide on the minimum benefit B_m acceptable. These values can then be substituted into the appropriate inequality in Table 2 and a determination can be made as to whether the remaining variable is likely to have a value which will satisfy that inequality so that at least the minimum benefit B_m can be attained.

As an example, consider a construction standard which would require the addition of hand-holds to boats so that a victim in the water holding onto a boat would be less likely to lose his grip. Suppose there is an annual average of 2021 accident victims who enter the water but hold onto their boats, 1714 of whom manage to remain with their boats and 307 of whom lose their grip.

---

* However, there are intricacies we have not yet discussed. Be certain to read Section 3.3.3.

TABLE 2. DETERMINING IF A MINIMUM BENEFIT CAN BE ACHIEVED

To test for a minimum benefit B_m, use either of the quantities W or Z in the appropriate inequality. If this inequality is satisfied, the minimum benefit is (theoretically) achievable, otherwise it is not.

Given an estimated α value, calculate:

$$W = \beta B_m + r \left((d - c)\alpha - (b - a)\beta \right) . \tag{3-15}$$

If $W > 0$, check the inequality

$$x \leq \frac{1}{W} \left[r(bc - ad) - B_m d \right] \tag{3-16}$$

If $W < 0$, check the inequality

$$x \geq \frac{1}{W} \left[r(bc - ad) - B_m d \right] \tag{3-17}$$

If $W = 0$, check the inequality

$$B_m \leq \frac{r}{d} (bc - ad) \tag{3-18}$$

Given an estimated x value, calculate:

$$Z = r(b + d - a - c) - B_m . \tag{3-19}$$

If $Z > 0$, check in inequality

$$\alpha \leq \frac{1}{xZ} \left[r(bc - ad + bx - ax) - B_m (d + x) \right] \tag{3-20}$$

If $Z < 0$, check the inequality

$$\alpha \geq \frac{1}{xZ} \left[r(bc - ad + bx - ax) - B_m (d + x) \right] \tag{3-21}$$

If $Z = 0$, check the inequality

$$x \leq \frac{r(bc - ad) - B_m d}{r(d - c)} \tag{3-22}$$

Suppose further that the survival probability of victims who do not lose their grip is 0.98, while the survival probability of those who lose their grip is only 0.43. Calculating the number of survivors in each case and presenting the quantities schematically, we have:

S_0	S_1
$\frac{a}{b} = \frac{132}{307} = 0.43$	$\frac{c}{d} = \frac{1680}{1714} = 0.98$

We estimate that 80% of the unreported victims are in states S_1, i.e., $\beta = 0.80$. Hence, α is estimated to be 0.20. It is estimated that $r = 0.30$, that is, that hand-holds would enable 30% of those victims who currently lose their grip to retain it. A minimum benefit B_m of ten lives saved is desired.

Using equation 3-15 in Table 2, we calculate:

$$W = (0.80)(10) + 0.30((1714 - 1680)(0.20) - (307 - 132)(0.80))$$
$$= -32.0 .$$

As $W < 0$ we check:

$$x \geq \frac{1}{W} \left[r(bc - ad) - B_m d \right]$$

$$x \geq \frac{1}{-32.0} \left[0.30 ((307)(1680) - (132)(1714)) - 10(1714) \right]$$

$$x \geq -2179 .$$

Obviously, any admissible value of x satifies this inequality, so it appears that the standard would meet the minimum benefit requirement of saving at least ten lives.

An alternative method involves constructing a table of benefit values for several α, x - combinations. The table might display α values along the top row, x values down the left column and B values in the body. The table could then be examined to see if reasonable values for α and x would yield acceptably high benefit values.

A third method would be to construct level curves * for selected benefit values and then, as with the table, determine if reasonable values for α and x would yield acceptably high benefits. Equation 3-14 or either of the following equivalent equations could be used to construct the curves:

$$\alpha = \frac{r(bc - ad + bx - ax) - B(d + x)}{x\ (r(b + d - a - c) - B)} \tag{3-23}$$

$$x = \frac{r(bc - ad) - Bd}{r\ \left[(b + d - a - c)\ \alpha - (b - a) \right] + (1 - \alpha)B}\ . \tag{3-24}$$

As an example of this method, consider the hand-hold example just presented. Retaining the same data on known victims we have, as before:

S_0	S_1
$\frac{a}{b} = \frac{132}{307} = 0.43$	$\frac{c}{d} = \frac{1680}{1714} = 0.98$

If we also retain our estimate of r = 0.30, then equation 3-23 (or 3-24) enables us to construct level curves for various benefit values. Figure 3-1 illustrates several such curves. Note that admissible α and x values fall in the ranges $0 \leq \alpha \leq 1$, $0 \leq x$. Thus, we see that if there are, say, at least 1000 unreported victims, then a benefit of about 50 lives saved is not unreasonable. This is of course based on an estimate of r = 0.30. Curves could be similarly drawn using different values of r.

In general, in using the level curve method, if a judgment is reached that acceptably high benefits would likely result from the contemplated regulation, a more careful analysis can be performed. In this analysis, more careful estimates of α , x and r could be made, and benefits could then be calculated using equation 3-14. This is discussed in the following pages.

---

* A level curve is a graph of all (α , x) points which yield the same value of B. That is, for each fixed value of B, equation 3-23 (or 3-24) gives a different level curve.

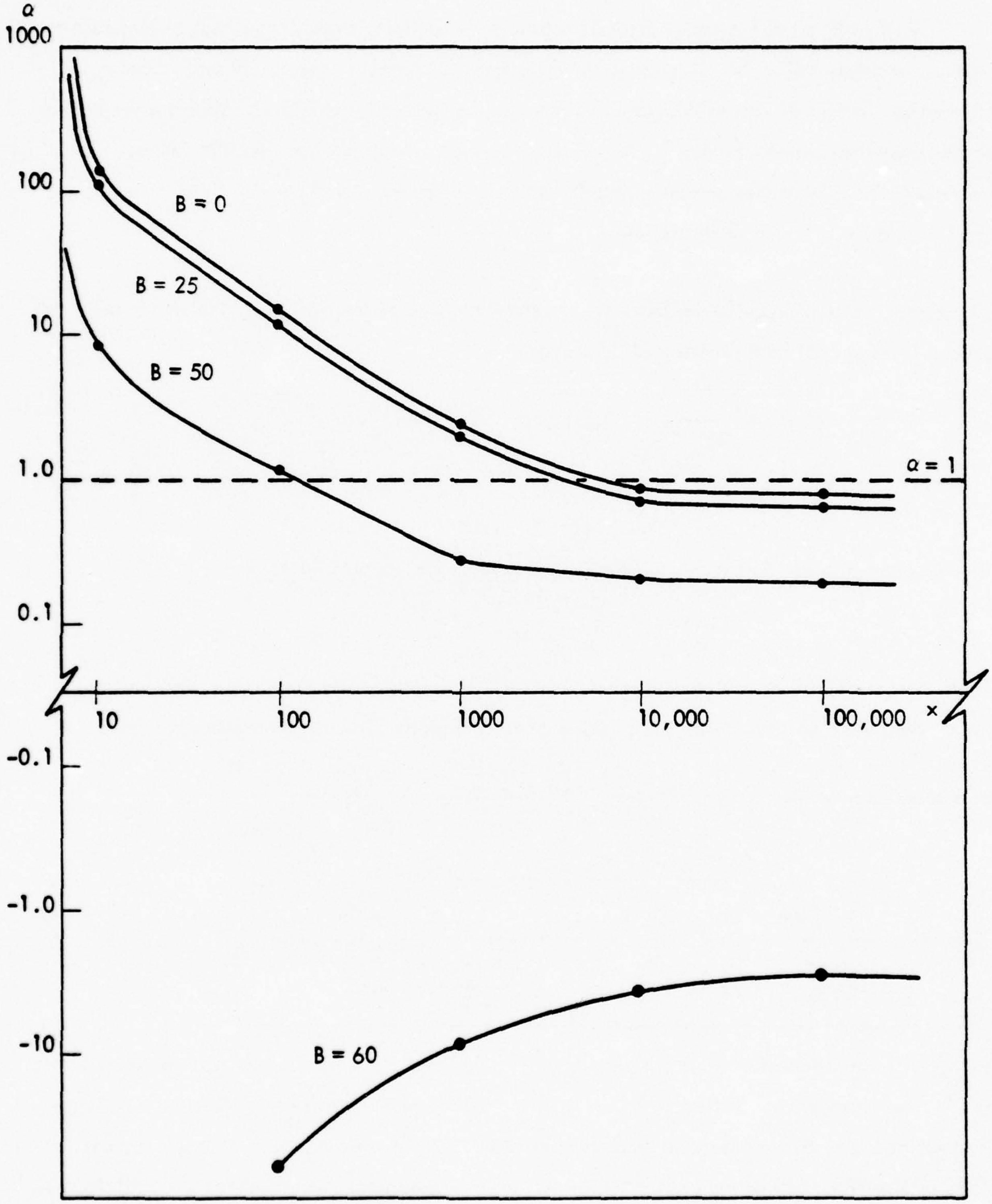


FIGURE 3-1. LEVEL CURVES IN AN EXAMPLE BENEFIT ESTIMATION PROBLEM

3.3.2.2 Error Bounds on Benefit Estimates — In many cases it may not be possible to estimate a single value for x or for α or for either of these variables. In such a case, by estimating a range of values that the variables might take it is possible to obtain error bounds on the expected benefit B; that is, to obtain a corresponding range of benefit values. In instances where estimates cannot be made with any degree of confidence, limiting processes can be used to arrive at error bounds.

Let us recall equation 3-14 which expresses the expected benefit of a regulation (in average annual lives saved) as a function of x and α :

$$B = B(\alpha, x) = \frac{r}{d + \beta x} \left(bc - ad + \left[(b - a)\beta - (d - c)\alpha\right] x\right) \tag{3-14}$$

where $\beta = 1 - \alpha$.

This formula may be used to calculate benefit values for any values of x and α , $0 \leq x$, $0 \leq \alpha \leq 1$.

Table 3 contains all of the inequalities and other information, including limiting values, which should be needed in the calculation of error bounds. We present some examples.

In a prior example involving PFD wear, the following data was used.

Reported Victims, Pre-Regulation

S_0 (non-wear)	S_1 (wear)
$\frac{a}{b} = \frac{3598}{4136} = 0.87 = p_0$	$\frac{c}{d} = \frac{2267}{2313} = 0.98 = p_1$

Using an estimated value of $\alpha = 0.85$, it was found that $p'_0 < p'_1$ so positive benefits would result from transferring victims from S_0 to S_1 , that is, from increasing PFD wear. Suppose it is estimated that a contemplated regulation would achieve an r value of 0.20; that is, 20% of the PFD non-wearers could be made into PFD wearers. To calculate the anticipated benefit B from such a regulation, it is necessary to know the number x of unreported accident victims in these states. Suppose, however, that we are unable to estimate this number. We can use Table 3 to obtain error bounds on B.

As no estimate for x is given, we use the values $x_1 = 0$, $x_2 = \infty$ as given in part B.2. of the table. Comparing α with $\frac{b}{b+d}$ we have:

$$\frac{b}{b+d} = \frac{4136}{4136+2313} = 0.64 < 0.85 = \alpha$$

So part A.2.b. of the table applies and we have:

$$B(0.85,\infty) \leq B(\alpha, x) \leq B(0.85,0) \ .$$

Using part B.2.b. of the table for $B(0.85,\infty)$ we have:

$$B(0.85,\infty) = \frac{0.20}{0.15}\left[(4136 - 3598)(0.15) - (2313 - 2267)(0.85)\right] = 55 \ .$$

Also, using part B.2.a. of the table,

$$B(0.85,0) = (0.20)(4136)(0.98 - 0.87) = 91 \ .$$

Thus, we obtain as error bounds:

$$55 \leq B \leq 91 \ .$$

So we can expect to save between 55 and 91 lives annually by promulgating the contemplated regulation.

As a second example, let us use the same data as in the first example, but replace our estimate of $\alpha = 0.85$ with a lower bound α_1 for α, say $\alpha_1 = 0.60$. Note that 0.60 satisfies condition A in Table 1 for positive benefits, but it is possible that some values of $\alpha > 0.60$ might yield negative benefits. We use Table 3 to determine what the benefit range is.

As we have no upper bound α_2 for α we take $\alpha_2 = 1$. Now $\frac{b}{b+d} = 0.64$, so $\alpha_1 \leq \frac{b}{b+d} \leq \alpha_2$ and part A.3.c. of Table 3 applies. Note that in this case only an upper bound x_2 for x is needed, and as we have no estimate of x_2 we use $x_2 = \infty$.

TABLE 3. BENEFIT ERROR BOUNDS RESULTING FROM UNREPORTED VICTIMS

Let $0 \leq \alpha_1 \leq \alpha \leq \alpha_2 \leq 1$, $0 \leq x_1 \leq x \leq x_2$. Then,

A. These basic inequalities may be used.

1. $B(\alpha_2, x) \leq B(\alpha, x) \leq B(\alpha_1, x)$ (3-25)

2(a). If $\alpha \leq \frac{b}{b+d}$, then $B(\alpha, x_1) \leq B(\alpha, x) \leq B(\alpha, x_2)$ (3-26)

2(b). If $\alpha \geq \frac{b}{b+d}$, then $B(\alpha, x_2) \leq B(\alpha, x) \leq B(\alpha, x_1)$ (3-27)

3(a). If $\alpha_2 \leq \frac{b}{b+d}$, then $B(\alpha_2, x_1) \leq B(\alpha, x) \leq B(\alpha_1, x_2)$ (3-28)

3(b). If $\frac{b}{b+d} \leq \alpha_1$, then $B(\alpha_2, x_2) \leq B(\alpha, x) \leq B(\alpha_1, x_1)$ (3-29)

3(c). If $\alpha_1 \leq \frac{b}{b+d} \leq \alpha_2$, then $B(\alpha_2, x_2) \leq B(\alpha, x) \leq B(\alpha_1, x_2)$. (3-30)

Note that the value x_1 is not used at all in 3(c); all that is needed is an upper bound x_2 on x.

B. If one or more bounds on α or x cannot be obtained, use the following in the appropriate inequality above.

1(a). $\alpha_1 = 0$

1(b). $\alpha_2 = 1$

2(a). $x_1 = 0$. $B(\alpha, 0) = B_0 = rb(p_1 - p_0) = \frac{r}{d}(bc - ad)$ (3-31)

2(b). $x_2 = \infty$. $B(\alpha, \infty) = \lim_{x_2 \to \infty} B(\alpha, x_2)$

$$= \frac{r}{\beta}\left[(b-a)\beta - (d-c)\alpha\right], \alpha < 1 \quad (3\text{-}32)$$

$$B(1, \infty) = -\infty, \text{ if } c < d$$

$$= r(b-a), \text{ if } c = d . \quad (3\text{-}33)$$

TABLE 3. BENEFIT ERROR BOUNDS RESULTING FROM UNREPORTED VICTIMS (concluded)

C. Irrespective of the error bounds derived above, the following always will hold.

1. If $p'_0 < p'_1$ for all (α, x) values under consideration, then $0 \leq B(\alpha, x)$ (3-34)

2. $B(\alpha, x) \leq r(b - a)$ (3-35)

D. Suppose the x unreported victims are distributed between states S_0 and S_1 in the same ratio as the reported survivors.

If $p_0 \leq p_1$, then $B_0 \leq B \leq \frac{1}{p_1} B_0$ (3-36)

and

if $p_0 \geq p_1$, then $\frac{1}{p_1} B_0 \leq B \leq B_0$, (3-37)

where $B_0 = rb(p_1 - p_0) = \frac{r}{d}(bc - ad)$ is the benefit and $p_0 = \frac{a}{b}$, $p_1 = \frac{c}{d}$ are the state S_0 and S_1 survival probabilities, all calculated by taking only reported victims into account.

Thus we have:

$$B(1,\infty) \leq B(\alpha,x) \leq B(0.60,\infty) .$$

Now using part B.2. of the table, as $c < d$,

$$B(1,\infty) = -\infty .$$

Also, $B(0.60,\infty) = \frac{0.20}{0.40} \left[(4136 - 3598)(0.40) - (2313 - 2267)(0.60)\right] = 94 .$

Thus, $-\infty \leq B \leq 94 .$

Obviously, these error bounds are too large. In order to obtain tighter bounds, we may resort to Table 1. From condition B of Table 1, we see that in order to have positive benefits when $\alpha = 1$, we must have:

$$x < \frac{(4136)(2267) - (3598)(2313)}{2313 - 2267} = 22,916 .$$

Unless we have good reason to believe that there are fewer than 22,916 unreported victims or we can place a more realistic upper bound on α, we must conclude that the contemplated regulation might result in a negative benefit — lives lost instead of saved.

As a final example, let us use the same numbers for reported victims and again take $r = 0.20$, but let us assume that the unreported victims act like the known survivors; that is, that the unreported victims are distributed between S_0 and S_1 in the same ratio as the known survivors. Then, from part D of Table 3,

$$B_0 \leq B \leq \frac{1}{p_1} B_0 .$$

Now, $B_0 = rb(p_1 - p_0) = (0.20)(4136)(0.98 - 0.87) = 91$

and $\frac{1}{p_1} B_0 = \left(\frac{1}{0.98}\right) (91) = 93 .$

Thus, we estimate that the contemplated regulation could result in saving from 91 to 93 lives per year, provided unreported victims act like known survivors.

As a postscript to this section we give a brief description of the chronology of the development of the techniques for dealing with the problem of unreported accidents. It has been well known to researchers in the area of recreational boating safety that while virtually all boating accidents involving fatalities are reported to the Coast Guard, most other boating accidents are not reported. Furthermore, much more detailed reports are made on fatal accidents than on non-fatal accidents. Thus, while the Coast Guard has some data on accident causes and recovery circumstances in fatal accidents it has much less of this data for non-fatal accidents.

It was realized that by considering only data from reported accidents in analyses of accident cause and victim recovery that distortions in results might occur. In particular, it was clear that in many, if not most, cases the recovery probabilities of accident victims in particular circumstances would be underestimated * . What would be the effect of such underestimation?

The case of estimating benefits by transferring victims from a state S_0 to a state S_1 was considered. In addition to the assumption that all fatalities are reported, a second assumption was initially made — that unreported victims have the same recovery circumstances as reported survivors. Under these assumptions, it was discovered that in most cases the difference $p'_1 - p'_0$ in recovery probabilities in the two states S_1 and S_0 would be less than that calculated by considering only data for reported victims * . It thus appeared that using probabilities derived only from data for reported victims would result in overestimation of benefits.

Further analysis, however, showed that the opposite was true. Under these assumptions, for $p_0 < p_1$ the inequality (3-36) $B_0 \leq B \leq \frac{1}{p_1} B_0$ was true, where B_0 was the benefit calculated by using only data for reported victims. So benefits would actually be under-estimated.

Later, equations such as 3-14 and inequalities such as 3-26 were developed giving benefits and error bounds under assumptions that non-reported victims were distributed between states in any pre-selected manner. Finally, methods, such as those involving Tables 1 and 2 were developed to help in determining if realistic choices for the variables α and x would likely result in positive benefits.

---

* The derivation of this result may be found in Appendix A.

3.3.3 Multi-State Benefit Analysis

Up to this point we have only discussed benefit estimation problems involving transferring victims from a single state S_0 to another state S_1. However, as the example below illustrates, considering transfer between only two states is too restrictive and often can lead to erroneous results. Although the data in the example is synthetic, it will illustrate precisely the type of problem that occurred when the Accident Recovery Model (Reference 1) was analyzed. The use of synthetic data enables us to present a simple example which clearly demonstrates the important principles involved. *

Suppose an educational program to increase PFD use is being considered. During an initial analysis of the Accident Recovery Model data base for PFD use, a startling discovery is made. The calculated recovery probability of accident victims in the water who do not use a PFD is found to be greater than the calculated recovery probability of victims in the water who do use a PFD. (This is the sort of unexpected difficulty reffered to in the first paragraph of Section 3.1.) Can it really be that PFD use is detrimental to survival, or is there another factor at work? Fortunately, it is the later explanation which applies. Not only is there an interaction between PFD use and survival, but there is also an interaction with other factors affecting both PFD use and survival. That is, survival probabilities and PFD use probabilities are conditional on other factors, so that, for instance, two victims in the state "PFD Used" need not have the same survuval probability. This is most easily seen with a numerical example.

Consider the following table of (synthetic) data which separates PFD use and non-use into two categories, adults and children. Each entry in the table is of the form:

$$\frac{\text{survivors}}{\text{victims}} = \text{survival probability} .$$

	Adults	Children
PFD Not Used	$\frac{980}{1000} = 0.98$	$\frac{200}{300} = 0.67$
PFD Used	$\frac{99}{100} = 0.99$	$\frac{250}{300} = 0.83$

* Reference 2 contains a more complex example using real data.

Note that PFD use is advantageous both for adults and for children; in both cases survival probabilities are greater with PFD use than without PFD use. However, when victims are not separated by age, the probability of survival for a victim not using a PFD is calculated as:

$$p_0 = \frac{980 + 200}{1000 + 300} = \frac{1180}{1300} = 0.91$$

while the survival probability for a victim using a PFD is calculated as:

$$p_1 = \frac{99 + 250}{100 + 300} = \frac{349}{400} = 0.87 \; .$$

The survival probability for PFD non-use is greater than for PFD use!

As will be seen later, this apparent anomaly is due to the distribution of PFD use among adults and children. It is a result of the fact that in the example adults have a greater survival probability than do children, but adults use PFDs much less frequently than do children.

In this example and in the general discussion which follows it one must remain cognizant of the cause of the problem. It is not a result of sample data being unrepresentative. Rather, it is a result of the actual circumstances in which accidents occur. The survival probability of a victim normally depends on many interrelated factors, and thus considering survival as a function of only a broad category, such as PFD wear, is actually an unrealistic simplification of more complex circumstances. Such unjustified simplification can lead to erroneous and even ridiculous results.

Several questions arise: What is the relationship between the overall survival probability in a state and the survival probabilities in sub-states where other factors are taken into account? How should benefits be calculated? When must other factors be taken into account and when can they be safety ignored?

While research into all aspects of the problem is far from complete, it is possible to give at least partial answers to these questions. In order to express these answers in a precise manner, it is necessary to introduce some additional notation.

Let S_0 and S_1 be two mutually exclusive * states and let $T_1, T_2, \ldots, T_n$ be a collection of mutually exclusive states. Let $S_0 T_i$, $1 \leq i \leq n$, denote that sub-state of S_0 and T_i which is determined by the defining conditions for both S_0 and T_i. That is, considering S_0 and T_i as sets of victims, $S_0 T_i$ is the set intersection $S_0 \cap T_i$, so that a victim is in state $S_0 T_i$ if and only if he is in both state S_0 and in state T_i. Let $S_1 T_i$ be defined similarly. For instance, if S_0 and S_1 denote PFD non-use and use, respectively, and T_1 and T_2 denote adults and children, respectively, then $S_0 T_1$ denotes adults not using PFDs, $S_1 T_2$ denotes children using PFDs, etc.

Also, for each i, $1 \leq i \leq n$, let a_i, b_i and p_{0i} denote the number of survivors, number of victims and survival probability, respectively, in state $S_0 T_i$, and let c_i, d_i and p_{1i} denote the number of survivors, victims and survival probability, respectively, in state $S_1 T_i$ **. Let a, b and p_0 denote the number of survivors, number of victims and survival probability, respectively, in state S_0, and let c, d and p_1 denote the number of survivors, number of victims and survival probability in state S_1. Let q_{0i} (or q_{1i}) denote the probability that a victim in state S_0 (or S_1) is also in state T_i.

These definitions immediately lead to the following results:

$$
\begin{aligned}
&\circ \quad a = a_1 + a_2 + \ldots + a_n = \sum_{i=1}^{n} a_i \\
&\circ \quad b = b_1 + b_2 + \ldots + b_n = \sum_{i=1}^{n} b_i \\
&\circ \quad c = c_1 + c_2 + \ldots + c_n = \sum_{i=1}^{n} c_i \\
&\circ \quad d = d_1 + d_2 + \ldots + d_n = \sum_{i=1}^{n} d_i
\end{aligned}
\tag{3-38}
$$

---

* By this we mean that a victim cannot be in more than one of the mutually exclusive states. Expressed in set-theoretic notation, $S_0 \cap S_1 = \emptyset$ and $T_i \cap T_j = \emptyset$ for $i \neq j$, $0 \leq i,j \leq n$.

** Although we use the notations p_0, p_1, etc., reserved earlier for statistics based on reported accidents and victims, in this section all quantities may be considered as referring either to statistics based on reported victims or to statistics based on all victims; the results derived hold in either case.

- $p_{0i} = \frac{a_i}{b_i}, \; 1 \leq i \leq n$
- $p_{1i} = \frac{c_i}{d_i}, \; 1 \leq i \leq n$ (3-39)

- $p_0 = \frac{a}{b}$
- $p_1 = \frac{c}{d}$ (3-40)

- $q_{0i} = \frac{b_i}{b}, \; 1 \leq i \leq n$
- $q_{1i} = \frac{d_i}{d}, \; 1 \leq i \leq n$ (3-41)

- $q_{01} + q_{02} + \ldots + q_{0n} = \sum_{i=1}^{n} q_{0i} = 1$
- $q_{11} + q_{12} + \ldots + q_{1n} = \sum_{i=1}^{n} q_{1i} = 1$ (3-42)

Since these definitions and results are a lot to absorb at one time, it will help to relate them to the above example involving PFD use in adults and children. We rewrite and expand the table in that example, including our new notation.

	T_1 (Adults)	T_2 (Children)	(All Victims)
S_0 (PFD not used)	$\frac{a_1}{b_1} = \frac{980}{1000} = 0.98 = p_{01}$	$\frac{a_2}{b_2} = \frac{200}{300} = 0.67 = p_{02}$	$\frac{a}{b} = \frac{1180}{1300} = 0.91 = p_0$
S_1 (PFD used)	$\frac{c_1}{d_1} = \frac{99}{100} = 0.99 = p_{11}$	$\frac{c_2}{d_2} = \frac{250}{300} = 0.83 = p_{12}$	$\frac{c}{d} = \frac{349}{400} = 0.87 = p_1$

Most of the quantities are presented in the table. The values for q_{0i} , q_{1i} are not. They are:

$$q_{01} = \frac{1000}{1000 + 300} = \frac{1000}{1300} = 0.77$$

$$q_{02} = \frac{300}{1000 + 300} = \frac{300}{1300} = 0.23$$

$$q_{11} = \frac{100}{100 + 300} = \frac{100}{400} = 0.25$$

$$q_{12} = \frac{300}{100 + 300} = \frac{300}{400} = 0.75$$

Now, the first of the questions we posed concerning the relationship between state and sub-state survival probabilities is answered by the following result.

o
$$p_0 = \frac{1}{b}\left(b_1 p_{01} + b_2 p_{02} + \ldots + b_n p_{0n}\right) = \frac{1}{b}\sum_{i=1}^{n} b_i p_{0i}$$
$$= p_{01} q_{01} + p_{02} q_{02} + \ldots + p_{0n} q_{0n} = \sum_{i=1}^{n} p_{0i} q_{0i} \qquad (3\text{-}43)$$

and

o
$$p_1 = \frac{1}{d}\left(d_1 p_{11} + d_2 p_{12} + \ldots + d_n p_{1n}\right) = \frac{1}{d}\sum_{i=1}^{n} d_i p_{1i}$$
$$= p_{11} q_{11} + p_{12} q_{12} + \ldots + p_{1n} q_{1n} = \sum_{i=1}^{n} p_{1i} q_{1i} \qquad (3\text{-}44)$$

What these equations state is that the overall recovery probabilities in states S_0 and S_1 are weighted sums of the recovery probabilities in the individual states $S_0\ T_i$ and $S_1\ T_i$, the weights being proportional to the number of victims in $S_0\ T_i$ and $S_1\ T_i$.

To illustrate the equations, we use the data in our example.

$$p_0 = p_{01} q_{01} + p_{02} q_{02} = (0.98)(0.77) + (0.67)(0.23) = 0.91$$

$$p_1 = p_{11} q_{11} + p_{12} q_{12} = (0.99)(0.25) + (0.83)(0.75) = 0.87$$

The results stated in equations 3-43 and 3-44 are exactly what we would expect: The difficulties are caused by corresponding weights for states S_0 and S_1 not being the same; that is, $q_{0i} \neq q_{1i}$. As a result, certain recovery probabilities p_{0i} get weighted more heavily in the calculation of p_0 than do the corresponding recovery probabilities p_{1i} in the calculation of p_1 , and conversely. This is how apparent anomalies can occur, such as having the overall recovery probability for victims using PFDs less than for victims not using PFDs.

We will not attempt to give a completely general answer to the second posed question on how to calculate benefits, but will consider instead a special case, namely one in which a contemplated regulation is designed to transfer accident victims from state S_0 to state S_1 , but will not transfer victims between any T-states. That is, we only will consider how benefits are to be calculated when victims are transferred from state $S_0 T_i$ to $S_1 T_i$, $1 \leq i \leq n$.

In such a case benefits may be calculated separately for each transfer, from $S_0 T_i$ to $S_1 T_i$, and the results summed to arrive at the overall expected benefit B. In particular, if r_i , $1 \leq i \leq n$, is the transfer rate (fraction) of victims transferred from $S_0 T_i$ to $S_1 T_i$, then the overall expected benefit is given by:

$$\begin{aligned} B &= (r_1 b_1)(p_{11} - p_{01}) + (r_2 b_2)(p_{12} - p_{02}) + \ldots + (r_n b_n)(p_{1n} - p_{0n}) \\ &= \sum_{i=1}^{n} (r_i b_i)(p_{1i} - p_{0i}) \end{aligned} \tag{3-45}$$

To illustrate with our example, suppose a contemplated educational program is designed to cause 30% of those adults who would not otherwise use a PFD in an accident to use one, and to cause 50% of the children who would not otherwise use a PFD to use one. Then $r_1 = 0.30$ and $r_2 = 0.50$ and the benefit of this program would be:

$$\begin{aligned} B &= (r_1 b_1)(p_{11} - p_{01}) + (r_2 b_2)(p_{12} - p_{02}) \\ &= (0.30)(1000)(0.99 - 0.98) + (0.50)(300)(0.83 - 0.67) \\ &= 27 \text{ lives saved .} \end{aligned}$$

We now turn to our third question concerning when other factors must be taken into account and when they may be safely ignored. While further research may yield additional criteria, the following two criteria furnish good guidelines to when states may be combined.

(a) If the recovery probabilities in states $S_0 T_i$ and $S_0 T_j$ are equal, the recovery probabilities in states $S_1 T_i$ and $S_1 T_j$ are equal and the transfer rates are the same, then states T_i and T_j may be combined.

(b) If the fraction of victims in state T_i who are in state S_0 is the same as the fraction of victims in state T_j who are in S_0, and the transfer rates are the same, then states T_i and T_j may be combined.

The mathematical equivalent of criterion (a) is:

o $$p_{0i} = p_{0j} \text{ , } p_{1i} = p_{1j} \text{ and } r_i = r_j \tag{3-46}$$

Criterion (b) may be expressed mathematically in the following three equivalent forms:

o $$\frac{q_{0i}}{q_{1i}} = \frac{q_{0j}}{q_{1j}} \text{ and } r_1 = r_j \tag{3-47}$$

o $$\frac{b_i}{b_i + d_i} = \frac{b_j}{b_j + d_j} \text{ and } r_i = r_j \tag{3-48}$$

o $$\frac{b_i}{d_i} = \frac{b_j}{d_j} \text{ and } r_i = r_j \tag{3-49}$$

Note that criterion (a) cannot be replaced with the more general condition $p_{1i} - p_{0i} = p_{1j} - p_{0j}$ and $r_i = r_j$. We illustrate this with an example. Let $r_i = r_j = 0.1$, and consider the data $a_i = 100$, $b_i = 200$, $c_i = 20$, $d_i = 30$, $a_j = 90$, $b_j = 120$, $c_j = 110$ and $d_j = 120$.

Now $p_{1i} - p_{0i} = \frac{20}{30} - \frac{100}{200} = 0.17$ and $p_{1j} - p_{0j} = \frac{110}{120} - \frac{90}{120} = 0.17$, so the more general condition is satisfied. Using 3-45 to calculate benefits, we have $B = (0.1)\ (200)\ (0.17) + (0.1)\ (120)\ (0.17) = 5.44$. But if we combine states and calculate benefits we obtain $B = (0.1)\ (320) \left(\frac{20 + 110}{30 + 120} - \frac{100 + 90}{200 + 120} \right) = 8.73$.

To illustrate criterion (b), we again consider a PFD example, but one including both age and sex. Suppose accident victims are divided by age, sex and PFD use as indicated in the following table.

	T_1 Male Adults	T_2 Female Adults	T_3 Male Children	T_4 Female Children
S_0 (PFD non-use)	$\frac{834}{850} = 0.98$	$\frac{146}{150} = 0.97$	$\frac{115}{167} = 0.69$	$\frac{63}{100} = 0.63$
S_1 (PFD use)	$\frac{69}{70} = 0.99$	$\frac{30}{30} = 1.00$	$\frac{158}{200} = 0.79$	$\frac{108}{120} = 0.90$

Suppose $r_1 = 0.20$, $r_2 = 0.40$, $r_3 = 0.60$, and $r_4 = 0.60$

We will illustrate that, because the second criterion applies (approximately) to states T_3 and T_4, they may be combined into the single state "children."

First, let us calculate benefits using all four states T_1, T_2, T_3, T_4:

$$\begin{aligned} B &= (0.20)\,(850)\,(0.99 - 0.98) + (0.40)\,(150)\,(1.00 - 0.97) + (0.60)\,(167)\,(0.79 - 0.69) \\ &\quad + (0.60)\,(100)\,(0.90 - 0.63) \\ &= 30 \text{ lives saved} . \end{aligned}$$

Now, we show that criterion (b) is met. Although it is only necessary to check one of equations 3-47, 3-48, and 3-49, we will check all three to illustrate the various values:

$$q_{03} = \frac{b_3}{b} = \frac{167}{1267} = 0.132$$

$$q_{13} = \frac{d_3}{d} = \frac{200}{420} = 0.476$$

$$q_{04} = \frac{b_4}{b} = \frac{100}{1267} = 0.0789$$

$$q_{14} = \frac{d_4}{d} = \frac{120}{420} = 0.286 .$$

Testing 3-47,

$$\frac{q_{03}}{q_{13}} = \frac{0.132}{0.476} = 0.277$$

and $$\frac{q_{04}}{q_{14}} = \frac{0.0789}{0.286} = 0.276 .$$

Since these values are approximately equal, the criterion is approximately satisfied and T_3 and T_4 may be combined.

Similarly, we could check equation 3-48, finding:

$$\frac{b_3}{b_3 + d_3} = \frac{167}{167 + 200} = 0.455$$

and $$\frac{b_4}{b_4 + d_4} = \frac{100}{100 + 120} = 0.455 \ .$$

Finally, to check equation 3-49, we would examine:

$$\frac{b_3}{d_3} = \frac{167}{200} = 0.835$$

and $$\frac{b_4}{d_4} = \frac{100}{120} = 0.833 \ ,$$

again obtaining approximate equality.

Combining states T_3 and T_4 we obtain the state T_3^* (children) and the following data table.

	T_1 (Male Adults)	T_2 (Female Adults)	T_3^* (Children)
S_0 (PFD non-use)	$\frac{834}{850} = 0.98$	$\frac{146}{150} = 0.97$	$\frac{178}{267} = 0.67$
S_1 (PFD use)	$\frac{69}{70} = 0.99$	$\frac{30}{30} = 1.00$	$\frac{266}{320} = 0.83$

Calculating benefits using this table, with $r_1 = 0.20$, $r_2 = 0.40$, and $r_3^* = 0.60$, we have:

$$B = (0.20)\ (850)\ (0.99 - 0.98) + (0.40)\ (150)\ (1.00 - 0.97) + (0.60)\ (267)\ (0.83 - 0.67)$$
$$= 29 \text{ lives saved} \ .$$

The one life difference between this calculation and the previous one is due to rounding error and the fact that the criterion was only approximately met.

The condition in criteria (a) and (b) that the transfer rates r_i and r_j be equal may be able to be relaxed in certain cases if a combined transfer rate r can be determined which has the same effect.

Of course, all of the above only partially answers the question of what factors need to be taken into account in multi-state benefit estimation. Realistically, all possible factors probably could not be checked against criteria (a) and (b). Furthermore, criterion (a) or (b) would very seldom, if ever, be more than approximately met. Finally, for some factors the states $S_1 T_i$, would contain no victims while the state $S_0 T_i$ would contain victims, making it necessary to assign a recovery probability p_{1i} on a judgmental basis. In order to choose the factors (states T_i) which would be most important to use in calculating benefits, some sort of test is necessary. One possibility is to use a Chi-Square test on all those factors which it is believed would be most likely to violate both criteria (a) and (b). Those factors (states) which test as having the most significant differences with respect to the equation terms in both criteria (a) and (b) would be the ones to use in benefit estimation.

Once the initial choice of the states T_i, $1 \leq i \leq n$, is made, it is advisable to use the same criteria to determine if it is necessary to further divide any of the states T_i. Of course, this could become a virtually endless process. Realistic decisions must be made and careful judgment used to decide when sufficient analysis has taken place. Careful assessment of the values of the recovery probabilities p_{0i} and p_{1i} are important. Certainly, if negative benefits are calculated in any transfer (a result of $p_{0k} > p_{1k}$), the affected state T_k should be carefully examined, for it may be that transfer in that state (i.e., from $S_0 T_k$ to $S_1 T_k$) is actually detrimental or it may be that the state T_k requires further subdivision. (We should mention at this point that the benefit estimation formulas included in earlier sections can be used in calculating negative as well as positive benefits.) If T_k contains relatively few victims it may be best to just omit it rather than attempt a subdivision which likely would lead to problems of insufficient data.

Once a final determination of the states T_i, $1 \leq i \leq n$, is made, benefits can be calculated for each transfer from $S_0 T_i$ to $S_1 T_i$. If the calculations have been made on the basis of data for reported victims, the methods of Section 3.3.2 may be used to take unreported victims into account in each $S_0 T_i$ to $S_1 T_i$ transfer. Finally, as in our PFD use examples, the separate benefits from each transfer are summed to obtain the overall benefit. If error bounds are used to account for unreported victims, the lower bounds and upper bounds should be summed separately to obtain the overall bounds.

In case the reader has missed the full significance of the problem discussed in this section, we point out some of its important aspects: Survival probabilities cannot be blindly used in performing benefit calculations. These probabilities are not always what they seem to be, because their values are affected by factors other than ones directly under study. Thus, it is possible for the overall survival probability in state S_0 to be greater than in state S_1, and yet for S_1 to actually be a more desirable state. Such was the case with the PFD use examples of this section. The reverse is also theoretically possible. It may be that the overall survival probability in state S_1 is greater than in state S_0 and yet S_1 is a less desirable state than is S_0. Again, it is <u>not</u> necessarily true that if $p_{1i} > p_{0i}$ for all i, $1 \leq i \leq n$, then $p_1 > p_0$.

In order to properly calculate benefits, it is necessary to carefully select states T_i in order to subdivide states S_0 and S_1 into sub-states of the form $S_0 T_i$ and $S_1 T_i$ and to compute benefits for transfers between these sub-states. If this subdivision is not performed in a careful, unbiased manner, an entire range of incorrect benefit estimates could result. The analyst must guard against playing the game of "*What states T_i should I choose* in order to achieve a reasonable (i.e., desired) answer?" While it is perfectly permissable and even desirable to judge the reasonableness of the result of a benefit calculation, such judgment should be a check. If this check indicates that the results are unreasonable, then further analysis of the problem can be performed. Hopefully, further research will result in better understanding of this problem and in additional means of dealing with it.

In closing this section, we mention the problem of calculating benefits for two or more regulations which may have interacting effects. Level flotation regulations and PFD education programs would be an example. One possible means of performing such benefit estimations would be to perform the calculations sequentially. For instance, the results of level flotation regulations could first be estimated and new victim data generated based on these estimates. Then, the effects of PFD educational programs could be calculated using the new victim data generated from the first benefit estimation. Sequential estimation in the reverse order could also be performed. Hopefully, the results would agree.

Even if only one regulation is being considered, the analyst must be careful to consider possible side effects of the regulation. For example, regulation to increase PFD use or availability might possibly result in reducing the number of victims remaining with their boats. Such side effects could result in unexpected negative benefits. Thus, it is important that all aspects of a regulation be considered.

3.4 Forecasting Benefit Growth

In the first phase of benefit prediction, the effects of a regulation on the factors associated with accident occurrence and/or victim survival are estimated. The second phase of benefit prediction uses the information developed in the first phase to make an estimate of what benefit (accidents prevented, lives saved, etc.) would result from the regulation if its full effect could be immediately felt. The third phase of benefit prediction adjusts the "full effect" estimation found in Phase 2 for time effects.

Phases 1 and 2 were discussed in Sections 3.2 and 3.3. This section is devoted to a discussion of Phase 3. In many respects, this is the most difficult area in benefit estimation and assessment. No truly good methods of long-range quantitative forecasting have yet been developed. Even intermediate range forecasts are often inaccurate. We need only look at the disagreements in expert economic opinion and at the number of instances in which such opinion has been proven wrong to realize that we should not expect many benefit forecasts to be very accurate.

In most cases, the forecasting problem in benefit prediction can be separated into two related components. The first component adjusts for the implementation rate of the regulation. The second component adjusts for changing boating patterns affecting accident causation.

Each of these components is a function of time and of factors, such as social and economic conditions, which themselves vary with time. Consequently, for any regulation we can express both the implementation rate component and the boating pattern component as a function of the single variable time.

If a new regulation requires a boating standard to be met, then the implementation rate will be a function of the sales of new boats and the retirement rate for old boats which are of the

type subject to the regulation. Data on sales and retirement can be gathered in advance and included in a supplementary, forecasting data base.

If a new regulation is actually an educational program, then the implementation rate will be based on research specific to that program. This research must determine the rate at which the programs' message reaches and significantly influences boaters.

The boating pattern component is much more difficult to determine. Perhaps some accidents such as falls overboard are directly proportional to the number of hours of boater exposure, but accidents such as two-boat collisions are probably not. Thus, the boating pattern component may be very difficult to obtain for some types of accidents. In such cases, a great deal of expert judgment will be required.

Some very preliminary research has been performed into applying forecasting techniques to CG-357 statistics. The results were far from encouraging. Further research into the application of forecasting (time series analysis) techniques is definitely needed. Among the techniques which should be investigated are the following:

- Linear (double) moving average
- Linear (double) exponential smoothing
- Time series decomposition
- Adaptive filtering
- Simple and multiple regression
- Census II
- Foran
- Econometric models
- Auto Regressive Moving Average (ARMA) processes including the Box Jenkins method
- Filters

The interested reader is referred to Reference 3 for a general discussion of many of these techniques.

Forecasting methods will also be required in the costing aspect of regulatory effectiveness research. Changes in boat purchasing patterns as well as changes in material and labor costs will require forecasting.

Once forecasting methods have been chosen, the necessary software should be made available for use with the data base model. Additionally, a supplementary data base containing data necessary for forecasting (such as economic data, boat sales, weather data, etc.) should be developed and integrated into the main data base.

The final result should be a model which can take the results of research on the expected influence a regulation will have on accident or fatality causative factors, and determine the future benefit of such a regulation as well as the costs.

4.0 BENEFIT ASSESSMENT

4.1 Introduction

Benefit assessment or tracking is the process of determining what benefits have accrued as the result of the promulgation of a regulation. Wyle's research into benefit assessment techniques concentrated on two general approaches:

- comparisons of pre-regulation accident data with post-regulation data, and
- comparisons based on post-regulation accident data alone.

These approaches are discussed in Sections 4.1 and 4.2.

Since there is no way of absolutely determining what would have happened if a regulation had not been promulgated, there is no way of absolutely proving that results obtained through benefit assessment methods are correct. For this reason, it is highly desirable to have available a number of different benefit assessment techniques based on different assumptions so that results can be obtained by different means. The closer the agreement between results obtained by different techniques, the more likely that the results are reliable.

Assessing benefits in lives saved or fatalities prevented is in itself a difficult task. The task of assessing benefits in terms of reductions in accidents, property damage and injuries is much more difficult as only a fraction of the non-fatal accidents is reported, although the reporting rate is increasing. One of the reasons for the increase is that increasingly many insurance companies are sending Coast Guard Boating Accident Report forms along with insurance claim forms to boating accident victims. At this time, it appears doubtful if sufficiently accurate estimates of the total number of non-fatal accidents and associated injuries and property damage can be made to enable the detection of benefits derived from any but the most effective regulations. Thus, benefit assessment analyses will probably be restricted to fatal accident data, at least in the near future.

4.2 Benefit Assessment Techniques Employing Pre- and Post-Regulation Data

Benefit assessment techniques which employ pre-regulation as well as post-regulation accident data may employ one or both of the following approaches:

- Using the pre-regulation accident data with, perhaps, pre- and post-regulation data on social, economic, weather, etc., patterns to project what the current accident pattern would have been had regulations not occurred. The benefit is then calculated as the difference between this projected loss and the actual loss.

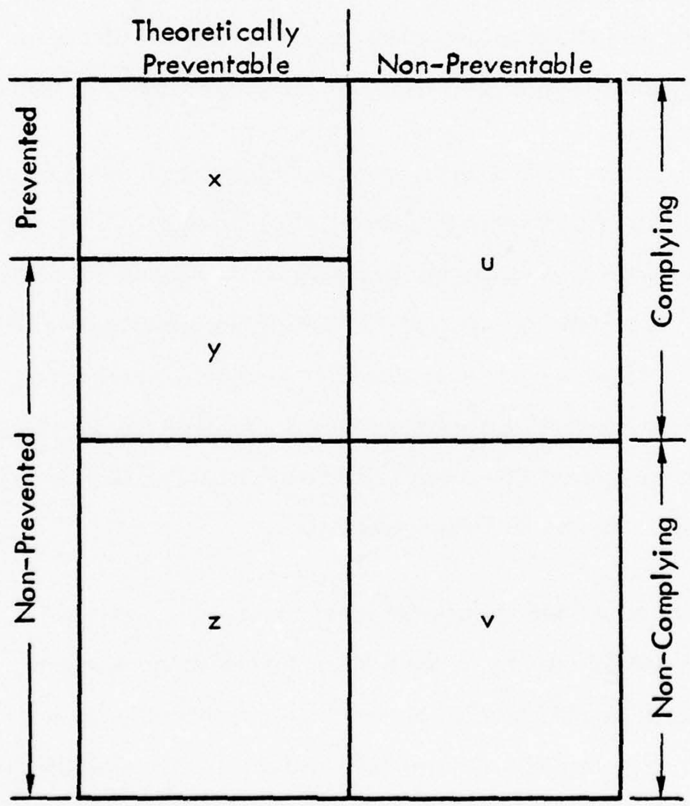
- Analyzing the pre- and post-regulation data for trends or changes, not readily explanable except as a result of the regulation

Regression analysis is often used to assess trends or to express one variable, say number of fatalities, as a function of another, say number of accidents. While regression analysis can be a powerful tool, it can also easily be misused. If it is used, correlation coefficients certainly should be calculated and statistical tests of the significance of the coefficients should be performed. However, even if such tests indicate significance, whenever possible the data should also be graphed so that the analyst can actually see the variation on the data. Appendix B contains an example of a regression analysis using real data which illustrates how easily regression results can be misinterpreted.

There are a number of quantitative forecasting techniques in addition to simple regression. The reader is referred to Reference 3 for a discussion of many of them. One of the areas of further research in both benefit prediction and assessment should be an in-depth investigation of the various forecasting techniques available so that a determination can be made as to which is most appropriate for application to boating accident data.

We now turn to a discussion of particular benefit assessment methods. The diagram below will be useful in both this and the following section. The diagram separates accidents (or fatalities, etc.) into classes on the basis of information obtained from the analysis of accident reports.

Suppose we are interested in assessing the benefit of a regulation promulgated in the past. Some vessels or boaters will be in compliance with the regulation, while others will not be. Furthermore, some accidents or fatalities should have been prevented as a result of compliance with the regulation. Others theoretically should have been prevented, but for some unknown reason were not prevented. Other accidents would have been theoretically preventable if non-complying boats had been in compliance. Finally, some accidents would not be preventable by compliance; that is, these accident occurrences are independent of compliance. The following Benefit Assessment Diagram separates accidents on the basis just described.



BENEFIT ASSESSMENT DIAGRAM

The variables represent numbers of accidents or potential accident situations (situations which would become accidents were it not for compliance with the provisions of the regulation) which occur during some specified period — one year, two years, etc. x represents the number of potential accident situations which, because of compliance with the regulation, did not become accidents. The value of x is unknown; no one reports non-accidents. The variable y is the number of accidents involving complying boats which should have been prevented by compliance, but which for some unknown reason were not prevented. z is the number of accidents which should have been prevented by compliance, but which involved non-complying boats. Finally, u and v are the numbers of accidents involving complying and non-complying boats which had accidents that would not have been affected by compliance; they would have occurred whether or not a boat complied with the regulation.

The determination of which category each accident is assigned to is made on the basis of a careful, structured analysis of the accident reports. This analysis should preferably be performed in such a manner that the analysts do not know if the involved boat(s) was in compliance until after a determination is made as to which classification, theoretically preventable or non-preventable, the accident belongs. The non-preventable accidents categories are not necessarily meant to include all types of non-preventable accidents, but rather only accidents which are related to at least some of the same factors as are the preventable accidents. For example, if a regulation is designed to affect only some collisions involving outboard boats less than sixteen feet (4.9 m) in length, then it may be preferrable to include only collisions involving this type of boat in the analysis. The determination of which accidents to include and which to exclude will depend on a careful assessment of the assumptions made in the benefit calculation and on the particular regulation being assessed.

Note that data is available on all of the variables except x. However, as discussed earlier, many non-fatal accidents are not reported so, at least in most instances, the variables will have to be restricted to fatal accidents.

We now describe one benefit estimation technique which makes use of the Benefit Assessment Diagram. Another technique making use of this diagram is described in Section 4.3.

Application of the Benefit Assessment Diagram: The Pre-Post Technique

The Pre-Post Technique is based on the assumption that, for the class of non-preventable accidents selected:

- changes over time in the number of theoretically preventable accident situations are proportional to changes in the number of non-preventable accident situations.

In other words, we wish to evaluate temporal changes in theoretically preventable accidents by measuring changes in non-preventable accidents.

The degree to which this assumption will be satisfied depends upon how carefully the class of non-preventable accidents is chosen and upon sample selection procedures.

The assumption can be applied to three classes of boats (or boaters):

(i) all sampled boats, both complying and non-complying,
(ii) only sampled complying boats
(iii) only sampled non-complying boats.

To describe the assumption in the form of equations, consider two benefit assessment diagrams, one for a pre-regulation accident period and one for a post-regulation accident period. Denote these diagrams by D' (pre-regulation) and D (post-regulation). Let x', y', z', u', v' be the variables of D' and x, y, z, u, v be the variables of D.

Let k, k' and k'' be the following quotients:

- (i) $k = \dfrac{x + y + z}{x' + y' + z'}$
- (ii) $k' = \dfrac{x + y}{x' + y'}$, if $x' + y' \neq 0$
- (iii) $k'' = \dfrac{z}{z'}$.

For each of the three boat classes listed above, the assumption states:

- (i) $k = \frac{u + v}{u' + v'}$
- (ii) $k' = \frac{u}{u'}$
- (iii) $k'' = \frac{v}{v'}$.

Note that these assumptions (equations) are not equivalent. It is possible to have some true and some false. Also we should expect in most cases to have $k'' < k < k'$ because the number of complying boats will normally increase relative to the number of non-complying boats after the regulation is promulgated.

Equating terms in the above equations and solving for the unknown variables x and x' yields:

$$\text{(i)} \quad x - kx' = \left(\frac{u + v}{u' + v'}\right) (y' + z') - (y + z) \tag{4-1}$$

$$\text{(ii)} \quad x - k'x' = \left(\frac{u}{u'}\right) y' - y \tag{4-2}$$

$$\text{(iii)} \quad \frac{z}{z'} = \frac{v}{v'} \ . \tag{4-3}$$

The third set of equations contains no unknown variables, but it may be used as a sort of test to help justify or refute the first two equations, and thus partially justify or refute the assumption at the beginning of this section upon which the Pre-Post Technique is based. This may be done using a Chi-Square test on the contingency table:

z	v	z + v
z'	v'	z' + v'
z + z'	v + v'	z + z' + v + v'

A very large value of χ^2 tends to indicate that the assumption is poorly satisfied, while a very small χ^2 value tends to indicate that the assumption is probably close to correct.

We may also compare the answers obtained using equations 4-1 and 4-2. If the answers are close, we have some justification in believing them to be valid. Answers obtained by the

method of Section 4.3 should also be compared to answers obtained using equations 4-1 and 4-2. The closer the agreement, the more likely the answers are correct.

Now, x' is the unknown number of accidents prevented in the pre-regulation period due to independent compliance with (what will become) the provisions of the regulation. The unknown terms kx' and $k'x'$ cannot be eliminated mathematically from equations 4-1 and 4-2. There are, however, three possible means of dealing with them:

(a) We might assume that the exposure change factor for actually preventable accident situations involving independently complying boats is also k. Under this assumption, the value of $x - kx'$ would be the number of accidents prevented due to the regulation itself and not due to independent compliance. We point out, however, that there very likely may be no good justification for making this assumption. (The same holds for $k'x'$.)

(b) The Benefit Assessment Diagram method described in Section 4.3 might be used to obtain a value for x'.

(c) We could eliminate from consideration all independently complying boats in the pre-regulation period, in effect arbitrarily setting $x' = y' = u' = 0$. Benefit calculations resulting from thie procedure would be based on the assumption that the proportion of boats in theoretically preventable accident situations would be the same for all boats after the regulation as it was for non-complying boats before the regulation. Mathematically,

$$\frac{x + y + z}{u + v} = \frac{z'}{v'} .$$

It should be noted that as $kx' \geq 0$ and $k'x' \geq 0$, the total benefit x due to compliance, either independently or as a result of the regulation, will be at least as large as the calculated values $x - kx'$ and $x - k'x'$.

We now turn to an example. A certain regulation was promulgated in 1971. The following table contains fatality data for two three-year periods, one ending two years prior to the year of promulgation, the other beginning two years after the year of promulgation. (Portions of this data are synthetic.)

Fatality Totals

	Theoretically Preventable		Non-Preventable	
	Complying	Non-Complying	Complying	Non-Complying
1973 - 1975	$y = 91$	$z = 170$	$u = 449$	$v = 193$
1967 - 1969	$y' = 159$	$z' = 262$	$u' = 266$	$v' = 289$

First, let us perform a contingency table test on equation 4-3, $\frac{z}{z'} = \frac{v}{v'}$:

170	193	363
262	289	551
432	482	914

For this table, $X^2 = 0.021$, a very small value, which encourages us to believe that the answers obtained using equations 4-1 and 4-2 will be close to correct.

Substituting the data into equation 4-1, we obtain:

$$x - kx' = \left(\frac{449 + 193}{266 + 289}\right) (159 + 262) - (91 + 170) = 226 .$$

Substituting the data into equation 4-2, we obtain:

$$x - k'x' = \left(\frac{449}{266}\right) (159) - 91 = 177 .$$

If we wish to make the assumption under (a), this tells us that about 200 lives were saved due to the regulation in the period 1973 - 1975. Even without this assumption it appears that at least 200 lives were saved during 1973 - 1975 as a result of independent or required compliance with the regulation.

One should not be disappointed that the results obtained from equations 4-1 and 4-2 do not agree more closely. It probably will be very difficult to obtain results from different calculations which do closely agree.

We will return to this example in the next section.

4.3 Benefit Assessment Techniques Employing Single Period Data

In Section 4.2 we derived equations for calculating benefit values based on an assumption concerning the relationship between pre- and post-regulation periods and preventable and non-preventable accidents. In this section we again use the Benefit Assessment Diagram to assess regulatory benefits. We will call the method employed the Compliance Technique. It is based on the following assumption:

- The fact of boat compliance or non-compliance does not affect the relative chance of a boat being in a theoretically preventable accident situation versus a non-preventable accident situation.

This assumption may be restated as:

- In any period, the odds of a boat (or boater) being in a theoretically preventable accident situation versus a non-preventable accident situation are the same for complying boats as for non-complying boats.

This assumption is much more likely to be satisfied in instances of equipment compliance rather than of boater action compliance. If boater decision-making is related to compliance, then it is likely that compliance is not independent of the presence of a potential accident situation, thus negating the assumption. This would occur, for instance, in the case of PFD use, where the use of PFDs would be related to boater characteristics and the environment in which a boater is located (see Section 3.3.3).

The assumption may be expressed as the equation:

- $\dfrac{x + y}{z} = \dfrac{u}{v}$

or equivalently,

- $x = \left(\dfrac{u}{v}\right) z - y$ (4-4)

As before, x is the total benefit due to compliance, whether the compliance is independent (voluntary) or is involuntary (directly due to legal requirements imposed by the regulation). The assumption should not be taken to mean that complying boats are assumed to be in imminent danger of having a theoretically preventable accident, but rather that they will be found operating in the same general conditions as non-complying boats.

As with the Pre-Post Technique, careful judgment has to be used in the choice of the non-preventable accident class and in the sample selection process. Also as before, during the accident analysis process, the determination of whether an accident is or is not theoretically preventable must be made independently of a knowledge of whether or not a boat is in compliance.

We illustrate this technique by applying it to the example of Section 4.2.

For the three-year, post-regulation period 1973-1975,

$$x = \left(\frac{u}{v}\right) z - y$$

$$= \left(\frac{449}{193}\right) (170) - 91$$

$$= 304 .$$

We may also apply the technique to the pre-regulation data:

$$x' = \left(\frac{u'}{v'}\right) z' - y'$$

$$= \left(\frac{266}{289}\right) (262) - 159$$

$$= 82 .$$

This value of x' may be substituted into the Section 4.2 results:

$$x - kx' = 226 \text{ and } x - k'x' = 177 .$$

Note that $k = \frac{u+v}{u'+v'} = \frac{642}{555} = 1.16$ and $k' = \frac{u}{u'} = \frac{449}{266} = 1.69$.

Making the substitutions we obtain:

$$x = 226 + kx' = 226 + (1.16)(82) = 321$$

and $$x = 177 + k'x' = 177 + (1.69)(82) = 316$$

Note that all of our results agree quite closely: Compliance, voluntary or involuntary, resulted in saving about 315 lives in the three year period 1973-1975.

A Chi-Square test can also be applied to the post-regulation accident data to determine if there is a significant difference in accident patterns of complying and non-complying boats.

The test is applied to the contingency table

y	u	y + u
z	v	z + v
y + z	u + v	y + z + u + v

Using this test and the post-regulation data of our example, we obtain the table

91	449	540
170	193	363
261	642	903

and a highly significant X^2 value of 93.5.

We should not expect to achieve this significant a result in most cases.

5.0 USERS GUIDE, CONCLUSIONS AND RECOMMENDATIONS

5.1 Users Guide

To help the individual who wishes to make use of the techniques developed in this report, the following user's guide is presented with steps outlined in the order in which they would most likely be followed. The guide covers only steps related to benefit prediction and assessment. Costing methods are not included.

It should be emphasized that the procedures and techniques presented in this report and outline below are preliminary in nature. They will be refined and/or replaced with better methods which will be developed in Phase II of this research program.

User's Guide

I. Pre-Regulation

 A. Develop data base model following guidelines of Section 2.0 (in order presented)

 B. Predict benefits

  1. Determine the causative factors (states and sub-states) to be affected by the regulation (Sections 3.1, 3.2, 3.3.3)

  2 Estimate the effects of the regulation on the accident or recovery causative factors; i.e., determine transfer rates between states and sub-states (Section 3.2)

  3. Make initial "full effect" benefit calculations using reported victim data (Section 3.3.1)

  4. Recompute "full effect" benefits taking into account unreported accidents and victims (Section 3.3.2)

   a. Check that taking unreported victims into account does not result in negative benefits (Table 1) or benefits below an acceptable minimum (Table 2) (Section 3.3.2.1)

   b. Estimate number of and distribution of unreported victims and calculate benefits; or calculate error bounds (ranges) of benefits based on range estimates of unreported victims (Table 3) (Section 3.3.2.2)

5. Combine benefits calculated for individual sub-state transfers (Section 3.3.3)

6. Forecast benefit growth taking time factor into account (Section 3.4)

II. Post-Regulation

A. Determine the benefit assessment techniques to be used (Section 4.0)

B. Update data base model including data required in the benefit assessment techniques chosen (Sections 2.0, 4.2)

C. Compute benefit assessments (Sections 4.2, 4.3)

D. Compare benefit assessments obtained by different methods (Sections 4.2, 4.3)

E. Compare assessed benefits with predicted ones

5.2 Conclusions and Recommendations

This report includes the results of initial research into the development of a methodology which the Coast Guard can use both for assessing (tracking) the results of its past actions and for predicting the results of proposed actions. As explained in the introduction, the initial research concentrated on benefit prediction and assessment. Future research will have to also include the development of methods for the prediction and tracking of costs.

The benefit estimation methodologies presented in this report should not be considered as sufficient for all estimation problems. For instance, they do not include any means for forecasting changes in boating and accident patterns, including, say, the effects of implementation rates of new construction standards. Research to determine appropriate forecasting techniques to be used in both benefit and cost prediction should be pursued. Certainly the problems associated with multi-state benefit estimation require further study so that reasonable guidelines in the use of the multi-state technique can be developed.

The reader should be aware that the benefit assessment techniques developed in this research do not directly evaluate the part played by boater decision-making vis-a-vi the existence of a regulation. It is important to consider the effect of boater decision-making in any case in which it is possible for such decision to negate the intent of a regulation. Boater decision-making can confound attempts at benefit assessment employing the usual techniques. The ORI

study (Reference 2) of bridge-to-bridge radiotelephone is probably an example. Ship operators may only have used bridge-to-bridge radiotelephones when they thought they would be of benefit. In terms of our Benefit Assessment Diagram this caused unaccounted for transfers from complying to non-complying status which made benefit assessment more difficult. Some initial research into specific aspects of this problem has been performed as part of the PFD task (Reference 2), where PFD use was separated into a base use rate plus an *environmental response use rate*. Additional research should be performed with the intent of developing more general benefit assessment techniques which can account for boater decision-making.

The methods developed in this project for assessing the benefits of past regulations are actually rather primitive. Further work to develop more sophisticated techniques is desirable.

In the area of costs it should be noted that while DOT has developed some costing methods in the area of automobile safety programs (Reference 4), many of these methods are not applicable to the small boat industry due to the differences in size and numbers of automobile and boat manufacturers. General methods should be developed for determining the cost effects of Coast Guard boating safety program on the boating industry (manufacturers, distributors, dealers, marinas, services) and on the boating public.

As boating activity patterns, accident patterns, cost patterns and Coast Guard regulatory actions are highly interrelated it is desirable to have future research consider all of these aspects and their interrelationships. Also, the Coast Guard is keenly aware that ill-considered regulations could have highly adverse effects on boaters and the boating industry. Consequently, further study should include research into the alternatives available and bounds on Coast Guard activities, taking into account its desire to promote safe boating without causing unnecessary, adverse effects. Such research will necessarily include a study of the relationships between possible Coast Guard activities and the patterns enumerated above. *In effect, it is suggested that a systems approach be taken in future regulatory effectiveness methodology research.*

REFERENCES

1. Personal Flotation Devices Research Phase I. Final Report. Prepared for the U.S. Coast Guard by Wyle Laboratories, August 1976

2. Spill-Risk Analysis Program Methodology Development and Demonstration. Vol I. Draft Final Report prepared for U.S. Coast Guard by Operations Research, Inc., 7 October 1975

3. Wheelwright, Steven C. and Spyros Makridakis. Forecasting Methods for Management. New York: John Wiley and Sons, Inc., 1973

4. Benefit and Cost Analysis Methodology-MVP Rule Making Programs. Report prepared for the National Highway Traffic Safety Administration by the Motor Vehicle Programs Engineering Systems Staff, August 1972.

5. Hays, William L. Statistics for the Social Sciences, (Second Edition). New York: Holt, Rinehart and Winston, Inc., 1973. p. 661

6. Dixon, Wilfred J. and Frank J. Massey, Jr., Introduction to Statistical Analysis, (Third Edition). New York: McGraw Hill Book Company, 1969

7. Crow, Edwin L., Frances A. Davis and Margaret W. Maxfield. Statistics Manual. New York: Dover Publications, Inc., 1960

APPENDIX A. DERIVATIONS OF BENEFIT PREDICTION RESULTS

In this appendix we derive the mathematical results presented in Section 3.0.

Suppose we have two disjoint states S_0 and S_1 with b average annual victims in S_0 and d average annual victims in S_1. Suppose a of the victims in state S_0 are survivors while c of the victims in state S_1 are survivors. Furthermore, let x, α and β be defined as follows:

- x is the average annual number of unreported accident victims in states S_0 and S_1 with α and β being the fractions of these victims in S_0 and S_1, respectively.

Note that αx is the number of unreported victims in state S_0, βx is the number of unreported victims in state S_1, and $\alpha + \beta = 1$.

Finally, suppose that it is estimated that when fully implemented a contemplated regulation will have the effect of transferring a fraction r of the victims in state S_0 to state S_1.

The situation for reported victims in states S_0 and S_1 may be illustrated schematically as follows:

S_0	S_1
$\frac{a}{b} = p_0$	$\frac{c}{d} = p_1$

where p_0 and p_1 are the recovery probabilities of victims in states S_0 and S_1, respectively.

Simple algebraic manipulations result in the following schematic for reported victims once the regulation has been fully implemented:

Reported Victims; Full Implementation of Regulation

$$\underset{S_0}{\boxed{\frac{(1-r)a}{(1-r)b} = p_0}} \xrightarrow{(rb)} \underset{S_1}{\boxed{\frac{c + p_1 rb}{d + rb} = p_1}}$$

This diagram indicates that (rb) victims are transferred to state S_1 leaving $(1 - r)b$ victims at state S_0. We assume:

- All victims at S_0 have an equal probability of being transferred to S_1.
- The recovery probability of a victim is determined solely by his state's defining criteria, i.e., there are no confounding interactions.

These assumptions guarantee that the probability of recovery at state S_0 remains p_0 after the transfer and that the transferred victims will have a new probability of recovery of p_1, so that the recovery probability at S_1 also remains unchanged.

With these assumptions it should be clear that $(1 - r)a$ survivors remain at S_0 and that the number of survivors at S_1 includes the original c survivors plus an additional $p_1 rb$ survivors.

In this simple model of benefit estimation we also assume that there are no changes in victim numbers other than that caused by the transfer.

Calculating for reported victims only, the benefit B_0 resulting from the regulation is the number of survivors saved after full implementation of the regulation less the number of pre-regulation survivors. Thus,

$$\begin{aligned} B_0 &= \left(a(1 - r) + c + p_1 rb\right) - (a + c) \\ &= a - ar + c + p_1 rb - a - c \\ &= p_1 rb - ar \\ &= r(p_1 b - a) \\ &= r(p_1 b - p_0 b) \\ &= rb(p_1 - p_0), \end{aligned}$$

and we have proven equation 3-2,

o $\quad B_0 = rb(p_1 - p_0)$.

To include unreported accident victims in our calculations, we make the assumption:

- All unreported victims are survivors.

With this assumption we can schematize the situation for all accident victims in S_0 and S_1 as follows:

All Victims; Without Regulation

S_0	S_1
$\frac{a + \alpha x}{b + \alpha x} = p'_0$	$\frac{c + \beta x}{d + \beta x} = p'_1$

All Victims; Regulation Fully Implemented

S_0	S_1
$\frac{(a + \alpha x)(1 - r)}{(b + \alpha x)(1 - r)} = p'_0$	$\frac{c + \beta x + p'_1 r (b + \alpha x)}{d + \beta x + r (b + \alpha x)} = p'_1$

The above expressions are obtained by merely making the appropriate substitutions ($a + \alpha x$ for a, etc.) in the corresponding expressions for reported victims.

Assuming a rate of transfer r, the benefit B of the regulation for all victims can be derived from 3-2 by making the same substitutions. Thus, we obtain equation 3-4,

o $\quad B = B(x, \alpha) = r(b + \alpha x)(p'_1 - p'_0)$.

The remaining results in Section 3.3 can be derived using a reexpressed form of this formula for B. Here are the steps in the derivation.

$$
\begin{aligned}
B &= r(b + \alpha x)(p'_1 - p'_0) \\
&= r(b + \alpha x)\left(\frac{c + \beta x}{d + \beta x} - \frac{a + \alpha x}{b + \alpha x}\right) \\
&= \frac{r(b + \alpha x)}{(d + \beta x)(b + \alpha x)}\Big((b + \alpha x)(c + \beta x) - (a + \alpha x)(d + \beta x)\Big)
\end{aligned}
$$

$$= \frac{r}{d + \beta x} \; (bc + b\beta x + c\alpha x + \alpha\beta x^2 - ad - a\beta x - d\alpha x - \alpha\beta x^2)$$

$$= \frac{r}{d + \beta x} \; (bc - ad + [\, b\beta - a\beta - d\alpha + c\alpha \,]\, x\,)$$

$$= \frac{r}{d + \beta x} \; (bc - ad + [\, (b - a)\beta - (d - c)\alpha \,]\, x\,)$$

This yeilds equation 3-14,

$$\text{o} \quad B = \frac{r}{d + \beta x} \; (bc - ad + [\, (b - a)\beta - (d - c)\,\alpha \,]\, x\,)\,.$$

Eliminating β by substituting $(1 - \alpha)$ we obtain another useful form,

$$\text{o} \quad B = \frac{r}{d + (1 - \alpha)x} \; (bc - ad + [\, b - a - (b + d - a - c)\,\alpha \,]\; x)\,. \qquad \text{(A-1)}$$

In manipulating inequalities it is necessary to consider whether factors are positive or negative. For reference, we now list some basic inequalities which are true, either because of the definitions of the quantities involved, or because of the assumptions which have been made.

$0 \leq a \leq b\,,\, b > 0$

$0 \leq c \leq d \quad d > 0$

$0 < r \leq 1$

$0 \leq \alpha \leq 1$

$0 \leq \beta \leq 1$

$x \geq 0$

We shall make use of these inequalities without specific reference.

To obtain the results in Table 1, we must find equivalent forms of the inequality $p'_0 < p'_1$. Because $B = r(b + \alpha x)\,(p'_1 - p'_0)$, $p'_0 < p'_1$ is equivalent to $B > 0$. Using equations 3-14 and A-1, this yields the equivalent inequalities:

$$bc - ad + [\, (b - a)\beta - (d - c)\alpha \,]\; x > 0 \qquad \text{(A-2)}$$

and $\quad bc - ad + [\, b - a - (b + d - a - c)\alpha \,]\, x > 0 \qquad \text{(A-3)}$

Now, the tests in part A of Table 1 are based on the assumption that $p_0 < p_1$, or equivalently that $ad < bc$. Thus, using this assumption, we have:

$$bc - ad > 0 \,,$$

and consequently, inequalities A-2 and A-3 will certainly be satisfied if:

$$(b - a)\beta - (d - c)\alpha \geq 0 \qquad \text{(A-4)}$$

or, equivalently,

$$b - a - (b + d - a - c)\alpha \geq 0 \,. \qquad \text{(A-5)}$$

Inequalities 3-7 is obviously equivalent to A-4, while 3-8 is clearly equivalent to A-5. Thus, if inequality 3-7 or 3-8 is satisfied, the condition $p'_0 < p'_1$ will be met.

If the two equivalent inequalities 3-7 and 3-8 are true (but $p_0 \geq p_1$, so condition A is not met), then:

$$(b - a)\beta - (d - c)\alpha \geq 0 \,.$$

If this inequality is strict, then we may divide inequality A-2 by $(b - a)\beta - (d - c)\alpha$ and obtain:

$$x > \frac{ad - bc}{(b - a)\beta - (d - c)\alpha} \,,$$

which is inequality (3-9).

Suppose, on the other hand, that inequalities 3-7 and 3-8 are not true. Then, $(b - a)\beta - (d - c)\alpha < 0$ and solving inequality A-2 for x we obtain:

$$x < \frac{ad - bc}{(b - a)\beta - (d - c)\alpha}$$

or
$$x < \frac{bc - ad}{(d - c)\alpha - (b - a)\beta}$$

which is inequality 3-10.

Finally, to obtain inequality 3-11, we solve inequality A-3 for α :

$$bc - ad + [\, b - a - (b + d - a - c)\alpha\,]\; x > 0$$

$$[\, b - a - (b + d - a - c)\alpha\,]\; x > ad - bc$$

$$b - a - (b + d - a - c)\alpha > (ad - bc)\,\frac{1}{x}$$

$$-(b - a + d - c)\alpha > (ad - bc)\,\frac{1}{x} - b + a$$

$$(b - a + d - c)\alpha < (bc - ad)\,\frac{1}{x} + b - a$$

$$\alpha < \frac{1}{(b - a) + (d - c)}\left((bc - ad)\,\frac{1}{x} + (b - a)\right).$$

As all of the inequalities in the above derivation are equivalent, 3-11 is equivalent to A-3, and thus $p'_0 < p'_1$ whenever 3-11 holds. Note that the derivation of 3-11 places no requirements on inequalities 3-7 or 3-8.

Equation 3-13 is the equation of the boundary curve of the region satisfying inequality 3-11. It may be derived from the equation $B = 0$, using the same derivation as that for inequality 3-11, but with the inequality symbols replaced by equality (=) symbols.

Derivations of the expressions in Table 2 are similar to those for Table 1. It is desired that the benefit B be greater than or equal to a minimum B_m ; that is, $B_m \leq B$. Using equations 3-14 and A-1, we have that $B_m \leq B$ is equivalent to both:

$$B_m \leq \frac{r}{d + \beta x}\,(bc - ad + [\,(b - a)\beta - (d - c)\alpha\,]\; x) \qquad \text{(A-6)}$$

and
$$B_m \leq \frac{r}{d + (1 - \alpha)x}\,(bc - ad + [\, b - a - (b + d - a - c)\alpha\,]\; x) \qquad \text{(A-7)}$$

We use inequality A-6 to derive the first expressions in Table 2.

Solving A-6 for an expression giving a bound on x, we obtain:

$$B_m (d + \beta x) \leq r(bc - ad) + r[\,(b - a)\beta - (d - c)\alpha\,]\, x$$

$$B_m d + \beta B_m x \leq r(bc - ad) + r[\,(b - a)\beta - (d - c)\alpha\,]\, x$$

$$\beta B_m x + r[\,(d - c)\alpha - (b - a)\beta\,]\, x \leq r(bc - ad) - B_m d$$

$$(\beta B_m + r[\,(d - c)\alpha - (b - a)\beta\,]\,)x \leq r(bc - ad) - B_m d$$

Now, if as in 3-15 we let:

$$W = \beta B_m + r[\,(d - c)\alpha - (b - a)\beta\,] \quad ,$$

then, the last inequality becomes:

$$Wx \leq r(bc - ad) - B_m d \; .$$

This expression is equivalent to A-6 and hence is equivalent to $B_m \leq B$. Furthermore, if $W > 0$ this expression is clearly equivalent to 3-16 while if $W < 0$ it is equivalent to 3-17. Finally, if $W = 0$, we have:

$$0 \leq r(bc - ad) - B_m d$$

or $$B_m \leq \frac{r}{d}(bc - ad)$$

which is inequality 3-18.

We have thus completed the derivation of the first half of Table 2. To derive the second half of this table, we use inequality A-7, solving it so as to obtain a bound on α:

$$B_m \leq \frac{r}{d + (1 - \alpha)x}(bc - ad + [\,b - a - (b + d - a - c)\alpha\,]\, x)$$

$$B_m d + B_m x - B_m x\alpha \leq r(bc - ad) + rx(b - a) - rx(b + d - a - c)\alpha$$

$$rx(b + d - a - c)\alpha - B_m x\alpha \leq r(bc - ad + bx - ax) - B_m d - B_m x$$

$$x[\,r(b + d - a - c) - B_m]\alpha \leq r(bc - ad + bx - ax) - B_m(d + x) \; .$$

Now, if as in 3-19 we let:

$$Z = r(b + d - a - c) - B_m ,$$

then the last inequality becomes:

$$xZ\alpha \leq r(bc - ad + bx - ax) - B_m(d + x)$$

or $$Z\alpha \leq \frac{1}{x} [r(bc - ad + bx - ax) - B_m(d + x)] .$$

This expression is equivalent to A-7 and hence is equivalent to $B_m \leq B$. It is now clear that inequality 3-20 is equivalent to $B_m \leq B$ when $Z > 0$, while inequality 3-21 is equivalent to $B_m \leq B$ when $Z < 0$. In case $Z = 0$, we have:

$$0 \leq \frac{1}{x} [r(bc - ad + bx - ax) - B_m(d + x)]$$

$$0 \leq r(bc - ad) + r(b - a)x - B_m d - B_m x$$

$$B_m x - r(b - a)x \leq r(bc - ad) - B_m d$$

$$\Big(B_m - r(b - a)\Big)x \leq r(bc - ad) - B_m d \qquad \text{(A-8)}$$

Now, as $Z = 0$,

$$0 = r(b + d - a - c) - B_m$$

so $$B_m = r(b - a) + r(d - c)$$

and $$B_m - r(b - a) = r(d - c) .$$

Thus, equation A-8 is equivalent to:

$$r(d - c)x \leq r(bc - ad) - B_m d$$

and $$x \leq \frac{r(bc - ad) - B_m d}{r(d - c)}$$

which shows inequality 3-22 is equivalent to $B_m \leq B$ in the case where $Z = 0$.

We turn now to equations 3-23 and 3-24. Equation 3-23 may be derived following exactly the same steps as were followed for inequality 3-20 with equality symbols replacing the inequality symbols, B replacing B_m, and the condition $Z > 0$ ignored.

Similarly, equation 3-24 can be derived in the same manner as inequality 3-16 with equality symbols replacing the inequality symbols, B replacing B_m, $(1 - \alpha)$ replacing β, and the condition $W > 0$ ignored.

To obtain the inequalities in Table 3, part A, is is first necessary to demonstrate the monotonicity of B, first as a function of α and then as a function of x.

Equation A-1 expresses B in the form suitable for differentiation with respect to α. Using this equation, we obtain:

$$\frac{\partial}{\partial \alpha} B(\alpha, x) = \frac{\partial}{\partial \alpha}\left[\frac{r}{d + (1-\alpha)x}\,(bc - ad + [\,b - a - (b + d - a - c)\alpha\,]\,x)\right]$$

$$= \frac{r}{(d + (1-\alpha)x)^2}\begin{bmatrix}(d + (1-\alpha)x)\,(b + d - a - c)\,(-x) \\ -\,(bc - ad + [\,b - a - (b + d - a - c)\alpha\,]\,x)\,(-x)\end{bmatrix}$$

$$= \frac{-rx}{(d + (1-\alpha)x)^2}\begin{bmatrix}bd + d^2 - ad - cd + bx + dx - ax - cx \\ -\,\alpha x(b + d - a - c) - bc + ad \\ -\,bx + ax + \alpha x(b + d - a - c)\end{bmatrix}$$

$$= \frac{-rx}{(d + (1-\alpha)x)^2}\,[\,bd + d^2 - cd + dx - cx - bc\,]$$

$$= \frac{-rx}{(d + (1-\alpha)x)^2}\,[\,b(d - c) + d(d - c) + x(d - c)\,]$$

$$= \frac{(-rx)\,(d - c)\,(b + d + x)}{(d + (1-\alpha)x)^2}$$

≤ 0 for all admissable values.

Thus, for any (fixed) value of x, $x \geq 0$, B is a decreasing* function of α.

---

* We adopt the usual convention of defining decreasing to mean strictly decreasing or constant.

We now use equation 3-14 to differentiate B with respect to x. To simplify the expressions, let $K = (b - a)\beta - (d - c)\alpha$. Equation 3-14 then becomes:

$$B = B(\alpha, x) = \frac{r}{d + \beta x} (bc - ad + Kx) .$$

Differentiating with respect to x we obtain:

$$\frac{\partial}{\partial x} B(\alpha, x) = \frac{\partial}{\partial x} \left[\frac{r}{d + \beta x} (bc - ad + Kx) \right]$$

$$= \frac{r}{(d + \beta x)^2} [(d + \beta x)K - \beta (bc - ad + Kx)]$$

$$= \frac{r}{(d + \beta x)^2} [dK + \beta Kx - \beta bc + \beta ad - \beta Kx]$$

$$= \frac{r}{(d + \beta x)^2} [dK - \beta bc + \beta ad]$$

As $\frac{r}{(d + \beta x)^2}$ is positive for all permissable values of x and α, we see that B is a monotonic function of x and that the sign of $(dK - \beta bc + \beta ad)$ will determine whether B is *increasing* * or decreasing. We therefore simplify this expression:

$$dK - \beta bc + \beta ad = d [(b - a)\beta - (d - c)\alpha] - \beta bc + \beta ad$$

$$= \beta bd - \beta ad - \alpha d^2 + \alpha cd - \beta bc + \beta ad$$

$$= \beta bd - \beta bc + \alpha cd - \alpha d^2$$

$$= \beta b(d - c) + \alpha d(c - d)$$

$$= (\beta b - \alpha d) (d - c) .$$

As $d - c \geq 0$, we see that:

B is an increasing function of x if and only if $\beta b - \alpha d \geq 0$

and B is a decreasing function of x if and only if $\beta b - \alpha d \leq 0$.

---

* Again, we adopt the usual convention, defining increasing to mean strictly increasing or constant.

Finally, note that $b\beta - \alpha d \geq 0$ is equivalent to:

$$\frac{\beta}{\alpha} \geq \frac{d}{b} \text{ and to } \alpha \leq \frac{b}{b+d} ,$$

while $\beta b - \alpha d \leq 0$ is equivalent to:

$$\frac{\beta}{\alpha} \leq \frac{d}{b} \text{ and to } \alpha \geq \frac{b}{b+d} .$$

Thus, for any (fixed) value of α ,

if $0 \leq \alpha \leq \frac{b}{b+d}$, B is an increasing function of x ,

and if $\alpha \geq \frac{b}{b+d}$, B is a decreasing function of x .

To derive the results of Table 3 we use the facts that for any $x \geq 0$, B is a monotonic decreasing function of α; for any $\alpha \leq \frac{b}{b+d}$, B is an increasing function of x; and for any $\alpha \geq \frac{b}{b+d}$, B is a decreasing function of x.

Let $0 \leq \alpha_1 \leq \alpha \leq \alpha_2 \leq 1, 0 \leq x_1 \leq x \leq x_2$. Results 3-25, 3-26 and 3-27 are merely restatements of the above monotonicity conditions.

To prove 3-28, suppose $\alpha_2 \leq \frac{b}{b+d}$. Then:

$$B(\alpha_2, x_1) \leq B(\alpha, x_1) ,$$

$$B(\alpha, x_1) \leq B(\alpha, x) \leq B(\alpha, x_2)$$

and $B(\alpha, x_2) \leq B(\alpha_1, x_2)$,

so $B(\alpha_2, x_1) \leq B(\alpha, x_1) \leq B(\alpha, x) \leq B(\alpha, x_2) \leq B(\alpha_1, x_2)$.

To prove 3-29, suppose $\alpha_1 \geq \frac{b}{b+d}$. Then:

$$B(\alpha_2, x_2) \leq B(\alpha, x_2) ,$$

$$B(\alpha, x_2) \leq B(\alpha, x) \leq B(\alpha, x_1)$$

and $B(\alpha, x_1) \leq B(\alpha_1, x_1)$,

so $B(\alpha_2, x_2) \leq B(\alpha, x_2) \leq B(\alpha, x) \leq B(\alpha, x_1) \leq B(\alpha_1, x_1)$.

To prove 3-30, suppose $\alpha_1 \leq \frac{b}{b+d} \leq \alpha_2$. Then:

$$B(\alpha_2, x_2) \leq B(\alpha_2, x)$$

$$B(\alpha_2, x) \leq B(\alpha, x) \leq B(\alpha_1, x)$$

and $$B(\alpha_1, x) \leq B(\alpha_1, x_2),$$

so $$B(\alpha_2, x_2) \leq B(\alpha_2, x) \leq B(\alpha, x) \leq B(\alpha_1, x) \leq B(\alpha_1, x_2).$$

Turning to parts B.1. and B.2(a) of Table 3, note that by definition, $0 \leq \alpha \leq 1$ and $x \geq 0$. Furthermore, it is clear from equation 3-14 that B is continuous for all (α, x), $0 \leq \alpha \leq 1$, $0 \leq x$. Thus, the choices $\alpha_1 = 0$, $\alpha_2 = 1$ and/or $x = 0$ are all appropriate when tighter bounds are unavailable.

The material in part B.2(b) is not actually needed, in the sense that we could always choose as a value for x_2 the number of boaters in the United States, or even the population of the United States. The limiting process does, however, yield simplified expressions and can yield quite tight bounds in some cases, as one of the examples illustrates. We therefore include it as an additional means of developing error bounds.

We define the notation $B(\alpha, \infty)$ as:

- $$B(\alpha,\infty) = \lim_{x \to \infty} B(\alpha, x), \; 0 \leq \alpha \leq 1.$$

Equations 3-14 can be used to evaluate this limit. We consider the following cases.

If $\alpha < 1$, then $\beta > 0$ and:

$$\begin{aligned}
B(\alpha,\infty) &= \lim_{x_2 \to \infty} \frac{r}{d + \beta x_2} \left(bc - ad + [(b-a)\beta - (d-c)\alpha]\, x_2\right) \\
&= \lim_{x_2 \to \infty} \frac{r(bc - ad)}{d + \beta x_2} + \lim_{x_2 \to \infty} \frac{r[(b-a)\beta - (d-c)\alpha]\, x_2}{d + \beta x_2} \\
&= 0 + r[(b-a)\beta - (d-c)\alpha] \lim_{x_2 \to \infty} \frac{1}{\frac{d}{x_2} + \beta} \\
&= \frac{r}{\beta}\left[(b-a)\beta - (d-c)\alpha\right]
\end{aligned}$$

If $\alpha = 1$, then $\beta = 0$ and:

$$B(1,\infty) = \lim_{x_2 \to \infty} \frac{r}{d} \left(bc - ad - (d - c)x_2\right) .$$

Now if $\alpha = 1$ and $c < d$, then $d - c > 0$ and:

$$B(1,\infty) = -\infty .$$

If $\alpha = 1$ and $c = d$, then:

$$\begin{aligned} B(1,\infty) &= \lim_{x_2 \to \infty} \frac{r}{d} \left(bd - ad - (0)x_2\right) \\ &= r(b - a) . \end{aligned}$$

This completes the derivation of the results of part B.2.b. However, as a point of interest, we also demonstrate that the expression:

$$B(\alpha,\infty) = \frac{r}{\beta} \left[(b - a)\beta - (d - c)\alpha \right]$$

can be used to derive $B(1,\infty)$, provided $B(1,\infty)$ is taken to mean:

$$B(1,\infty) = \lim_{\alpha \to 1^-} \beta(\alpha,\infty) .$$

Calculating this limit, we have:

$$\begin{aligned} B(1,\infty) &= \lim_{\alpha \to 1^-} \frac{r}{\beta} \left[(b - a)\beta - (d - c)\alpha \right] \\ &= r(b - a) - \lim_{\alpha \to 1^-} (d - c) \left(\frac{\alpha}{1 - \alpha} \right) . \end{aligned}$$

If $c < d$, then $d - c > 0$, and as $0 < \alpha < 1$, $B(1,\infty) = -\infty$

If $c = d$, then:

$$\begin{aligned} \beta(1,\infty) &= r(b - a) - \lim_{\alpha \to 1^-} (0) \left(\frac{\alpha}{1 - \alpha} \right) \\ &= r(b - a) . \end{aligned}$$

The demonstration is now complete.

Part C of Table 3 is derived as follows. The benefit B is given by equation 3-4:

$$B = B(\alpha, x) = r(b + \alpha x)(p'_1 - p'_0) .$$

If $p'_1 \geq p'_0$, this benefit is clearly non-negative which is the result in C.1.

The result in C.2. can be obtained intuitively or mathematically. Intuitively, the largest number of lives that could possibly be saved under our assumptions is the number of victims who are transferred to state S_1 who would perish if they remained at S_0. This number is $r(b - a)$.

Mathematically, the largest benefit will occur when $p'_1 = 1$. The benefit in this case is:

$$\begin{aligned} B &= r(b + \alpha x)(1 - p'_0) \\ &= r(b + \alpha x)\left(1 - \frac{a + \alpha x}{b + \alpha x}\right) \\ &= r(b + \alpha x - a - \alpha x) \\ &= r(b - a) . \end{aligned}$$

Again, we obtain result 3-35 in part C.2.

To obtain the result in part D of Table 3, a mathematical expression equivalent to the assumption that the x unreported victims are distributed between states S_0 and S_1 in the same ratio as the reported survivors is needed. For $\alpha < 1$, the expression:

$$\frac{\alpha}{\beta} = \frac{a}{c}$$

may be used. Solving this expression for α we have:

$$\begin{aligned} \frac{\alpha}{1 - \alpha} &= \frac{a}{c} \\ \alpha c &= a - a\alpha \\ \alpha &= \frac{a}{a + c} \end{aligned}$$

We will use part A.2. of Table 3 to obtain the bounds. Consequently, it is necessary to compare α with $\frac{b}{b+d}$.

If $p_0 \leq p_1$, then:

$$\frac{a}{b} \leq \frac{c}{d} ,$$

and $\frac{a}{c} \leq \frac{b}{d}$.

So, $\frac{a}{c} + 1 \leq \frac{b}{d} + 1$

and $\alpha = \frac{a+c}{c} \leq \frac{b+d}{d}$.

Similarly, if $p_0 \geq p_1$, $\alpha \geq \frac{b+d}{d}$.

Using the bounds $0 < x < \infty$, we have:

if $p_0 \leq p_1$, then $B(\alpha,0) \leq B(\alpha,x) \leq B(\alpha,\infty)$;

if $p_0 \geq p_1$, then $B(\alpha,\infty) \leq B(\alpha,x) \leq B(\alpha,0)$.

Now $B(\alpha,0) = B_0$. To obtain the bound at infinity, first consider the case where $\alpha < 1$. Using equation 3-32,

$$\begin{aligned} B(\alpha,\infty) &= \frac{r}{\beta}\ [\ (b-a)\beta - (d-c)\alpha] \\ &= r\left[(b-a) - (d-c)\frac{\alpha}{\beta}\right] \\ &= r\left[(b-a) - (d-c)\frac{a}{c}\right] \\ &= \frac{r}{c}\ (bc - ac - ad + ac) \\ &= \frac{r}{c}\ (bc - ad) \\ &= \frac{rbd}{c}\left(\frac{bc}{bd} - \frac{ad}{bd}\right) \end{aligned}$$

$$= \frac{d}{c}\ (rb)\left(\frac{c}{d} - \frac{a}{b}\right)$$

$$= \frac{1}{p_1}\ (rb)\ (p_1 - p_0)$$

$$= \frac{1}{p_1}\ B_0\ .$$

For the case $\alpha = 1$, all unreported victims are at state S_0. As the unreported victims are distributed in the same manner as the reported survivors, this means that there are survivors at S_0 and there are no survivors at S_1; that is, $a > 0$ and $c = 0$. Consequently, $p_0 > 0$ and $p_1 = 0$. This means $p_0 > p_1$ and therefore $B_0 = rb(p_1 - p_0) < 0$. We may therefore adopt the convention $\frac{1}{p_1} B_0 = \frac{1}{0} B_0 = -\infty$, for $\alpha = 1$.

Continuing with the case $\alpha = 1$, note that as $d > 0$, $c < d$. Hence, using equation 3-33 in Table 3, $B(\alpha, \infty) = B(1, \infty) = -\infty$. Thus, in the case $\alpha = 1$, we may write:

$$B(\alpha, \infty) = \frac{1}{p_1} B_0\ .$$

This equation was proven above for $\alpha < 1$, so we now have that for any α - value ($0 \leq \alpha \leq 1$) obtained under our assumption,

$$B(\alpha, \infty) = \frac{1}{p_1} B_0\ .$$

Combining our results, we obtain inequalities 3-36 and 3-37:

$$\text{if } p_0 \leq p_1\text{, then } B_0 \leq B \leq \frac{1}{p_1} B_0\ ,$$

$$\text{and if } p_0 \geq p_1\text{, then } \frac{1}{p_1} B_0 \leq B \leq B_0\ .$$

The only results stated in the postscript to Section 3.3.2.2 which have not already been proven are the intuitively obvious result that including unreported victims (all survivors) in the calculation of recovery probabilities will cause the probabilities to increase, and the less intuitive result that in most cases the difference in such probabilities will decrease.

The first of these results is equivalent to the inequality $\frac{a}{b} \leq \frac{a+y}{b+y}$, where a and b are as before and y is some number of unreported victims, all survivors. This can be proven as follows.

From the definitions, $a \le b$; therefore, as $y \ge 0$, $ay \le by$,

so $\quad ab + ay \le ab + by$,

$$a(b + y) \le b(a + y)$$

and finally,

$$\frac{a}{b} \le \frac{a + y}{b + y} .$$

While the second result is not true in all circumstances, it is true under fairly general conditions. We will prove it for the case when all survival probabilities are at least one-half. The result may be expressed mathematically in the notation we have been using as:

$$\text{if } p_0 \ge \frac{1}{2} \text{ and } \frac{\alpha}{\beta} = \frac{a}{c}, \text{ then } p'_1 - p'_0 \le p_1 - p_0 .$$

Now when

$$\frac{\alpha}{\beta} = \frac{a}{c}, \text{ then } \alpha = \frac{a}{a + c} \text{ and } \beta = \frac{c}{a + c},$$

and the inequality $p'_1 - p'_0 \le p_1 - p_0$ may be expressed as:

$$\frac{c + \dfrac{cx}{a + c}}{d + \dfrac{cx}{a + c}} - \frac{a + \dfrac{ax}{a + c}}{b + \dfrac{ax}{a + c}} \le \frac{c}{d} - \frac{a}{b} . \qquad \text{(A-9)}$$

We will prove this inequality.

For convenience, let $m = \dfrac{x}{a + c}$. The above inequality then becomes:

$$\frac{c + mc}{d + mc} - \frac{a + ma}{b + ma} \le \frac{c}{d} - \frac{a}{b} .$$

As all of the denominators involved are positive, we may "multiply them out" and simplify, obtaining the equivalent inequality:

$$abd^2 - a^2d^2 - ma^2cd - b^2cd + b^2c^2 + mabc^2 \ge 0 .$$

Factoring, we obtain another equivalent inequality:

$$(bc - ad)\,[\; bc + ad + mac - bd \;] \geq 0, \qquad \text{(A-10)}$$

but as $p_1 \geq p_0$,

$$\frac{c}{d} \geq \frac{a}{b}, \text{ and } bc - ad \geq 0 .$$

If $bc - ad = 0$, then inequality A-10 is trivially satisfied and our desired conclusion, A-9, is true.

If $bc - ad > 0$, then inequality A-10 is equivalent to $bc + ad + mac - bd \geq 0$.

Thus, we have that for $bc - ad > 0$, inequality A-9 is equivalent to:

$$bc + ad + mac - bd \geq 0 . \qquad \text{(A-11)}$$

Assuming $p_0 = \frac{a}{b} \geq \frac{1}{2}$, so that $2a \geq b$, and using $bc > ad$ we obtain:

$$\begin{aligned} bc + ad + mac - bd &> bc + ad - bd \\ &> ad + ad - bd \\ &= (2a - b)\, d \\ &\geq 0 \end{aligned}$$

Thus, A-11 is true and hence inequality A-9 is also true. This completes the proof.

Although the condition $\frac{a}{b} \geq \frac{1}{2}$ is a sufficient condition for the result just proven to hold, it is not a necessary condition. Indeed, as we saw in the proof, $bc + ad + mac - bd \geq 0$ is a necessary and sufficient condition provided $\frac{a}{b} < \frac{c}{d}$. This condition need not always hold as the example below illustrates.

Suppose that the number of reported survivors at state S_0 is 10 while the total number of reported victims at S_0 is 100. Suppose that the corresponding numbers at state S_1 are 10 and 50. The probabilities of recovery at these nodes are:

$$p_0 = \frac{10}{100} = 0.1 \qquad p_1 = \frac{10}{50} = 0.2 \ .$$

If there are an additional 200 unreported, recovered victims who are not taken into account but who should be distributed between these nodes in the same ratio as the reported survivors, then the revised recovery probabilities are:

$$p'_0 = \frac{110}{200} = 0.55 \qquad p'_1 = \frac{110}{150} = 0.733 \ .$$

In this case, the recovery probability at state S_1 has increased more than at state S_0, so:

$$p'_1 - p'_0 > p_1 - p_0 \ .$$

In Section 3.3.3, equations 3-38 through 3-42 are immediate consequences of the stated definitions. Equations 3-43 are also easily derivable:

$$p_0 = \frac{a}{b} = \frac{1}{b} \sum_{i=1}^{n} a_i = \frac{1}{b} \sum_{i=1}^{n} b_i \frac{a_i}{b_i} = \frac{1}{b} \sum_{i=1}^{n} b_i \ p_{0i}$$

and $$p_0 = \sum_{i=1}^{n} \frac{b_i}{b} p_{0i} = \sum_{i=1}^{n} p_{0i} \, q_{0i} \ .$$

Equations 3-44 are derived in exactly the same manner.

Each term $(r_i b_i)(p_{1i} - p_{0i})$ of equation 3-45 is merely an instance of equation 3-2 for calculating the benefit B_i resulting from a transfer between states $S_0 T_i$ and $S_1 T_i$. Because S_0 and S_1 are mutually exclusive and the states $T_1, \ldots, T_n$ are mutually exclusive, no victim can be in more than one of the states $S_0 T_1$, $S_0 T_2$, ..., $S_0 T_n$, $S_1 T_1$, $S_1 T_2$, ..., $S_1 T_n$. As a result, the benefits B_i may be summed to obtain the overall benefit B. Thus, we have equation 3-45:

$$B = \sum_{i=1}^{n} B_i = \sum_{i=1}^{n} (r_i b_i)(p_{1i} - p_{0i}) \ .$$

To validate criteria (a) and (b) it is necessary to prove that under either condition:

$$(r_i b_i)(p_{1i} - p_{0i}) + (r_j b_j)(p_{1j} - p_{0j}) = r(b_i + b_j)\left(\frac{c_i + c_j}{d_i + d_j} - \frac{a_i + a_j}{b_i + b_j}\right) \quad \text{(A-12)}$$

where r is the common value $r = r_i = r_j$,

$\dfrac{c_i + c_j}{d_i + d_j}$ is the survival probability of the combined states $S_1 T_i$ and $S_1 T_j$ (i.e., the state $S_1 \cap (T_i \cup T_j)$), and

$\dfrac{a_i + a_j}{b_i + b_j}$ is the survival probability of the combined states $S_0 T_i$ and $S_0 T_j$ (i.e., the state $S_0 \cap (T_i \cup T_j)$).

Substituting for p_{0i}, p_{1i}, p_{0j}, and p_{1j}, we obtain an equation equivalent to A-12:

$$(r_i b_i)\left(\frac{c_i}{d_i} - \frac{a_i}{b_i}\right) + (r_j b_j)\left(\frac{c_j}{d_j} - \frac{a_j}{b_j}\right) = r(b_i + b_j)\left(\frac{c_i + c_j}{d_i + d_j} - \frac{a_i + a_j}{b_i + b_j}\right) \quad \text{(A-13)}$$

We will use equation A-13 in validating the criteria.

First, suppose that criterion (a) is met. Then equations 3-46 hold; that is,

$$p_{0i} = p_{0j},\ p_{1i} = p_{1j} \text{ and } r_i = r_j = r.$$

Substituting for the probabilities, we have:

$$\frac{a_i}{b_i} = \frac{a_j}{b_j} \text{ and } \frac{c_i}{d_i} = \frac{c_j}{d_j}.$$

Therefore,

$$a_j = a_i\left(\frac{b_j}{b_i}\right) \text{ and } c_j = c_i\left(\frac{d_j}{d_i}\right).$$

Consequently,

$$r(b_i + b_j)\left(\frac{c_i + c_j}{d_i + d_j} - \frac{a_i + a_j}{b_i + b_j}\right) = r(b_i + b_j)\left(\frac{c_i + c_i\left(\frac{d_j}{d_i}\right)}{d_i + d_j} - \frac{a_i + a_i\left(\frac{b_j}{b_i}\right)}{b_i + b_j}\right)$$

$$= r(b_i + b_j)\left(\frac{\frac{c_i}{d_i}\;(d_i + d_j)}{d_i + d_j} - \frac{\frac{a_i}{b_i}\;(b_i + b_j)}{b_i + b_j}\right)$$

$$= r(b_i + b_j)\left(\frac{c_i}{d_i} - \frac{a_i}{b_i}\right)$$

$$= rb_i\left(\frac{c_i}{d_i} - \frac{a_i}{b_i}\right) + rb_j\left(\frac{c_i}{d_i} - \frac{a_i}{b_i}\right)$$

$$= (r_i\, b_i)\left(\frac{c_i}{d_i} - \frac{a_i}{b_i}\right) + (r_j\, b_j)\left(\frac{c_j}{d_j} - \frac{a_j}{b_j}\right)$$

and equation A-13 is true. Thus, if criterion (a) is met, the states T_i and T_j may be combined.

Turning to criterion (b), the statement of this criterion is the same as equation 3-48. We show all three equations, 3-47, 3-48 and 3-49 are equivalent.

Now, it is easy to see that the following equations are all *equivalent:*

$$\frac{b_i}{b_i + d_i} = \frac{b_j}{b_j + d_j} \tag{3-48}$$

$$b_i\,(b_j + d_j) = b_j\,(b_i + d_i)$$

$$b_i\, b_j + b_i\, d_j = b_i\, b_j + b_j\, d_i$$

$$b_i\, d_j = b_j\, d_i$$

$$\frac{b_i}{d_i} = \frac{b_j}{d_j} \tag{3-49}$$

$$\frac{\frac{b_i}{b}}{\frac{d_i}{d}} = \frac{\frac{b_j}{b}}{\frac{d_j}{d}}$$

$$\frac{q_{0i}}{q_{1i}} = \frac{q_{0j}}{q_{1j}} \tag{3-47}$$

Thus, we see that equations 3-47, 3-48, and 3-49 are equivalent.

Now, to prove the validity of criterion (b), suppose it is met; that is, suppose equations 3-47, 3-48, and 3-49 are satisfied. Let $r = r_i = r_j$.

As 3-49 is true,

$$b_i d_j = b_j d_i$$

so $$b_i d_i + b_i d_j = b_i d_i + b_j d_i ,$$

$$b_i (d_i + d_j) = (b_i + b_j) d_i$$

and thus,

$$\frac{b_i}{d_i} = \frac{b_i + b_j}{d_i + d_j} \tag{A-14}$$

We now prove equation A-13:

$$(r_i b_i)\left(\frac{c_i}{d_i} - \frac{a_i}{b_i}\right) + (r_j b_j)\left(\frac{c_j}{d_j} - \frac{a_j}{b_j}\right)$$

$$= r\left(\frac{c_i b_i}{d_i} - a_i\right) + r\left(\frac{c_j b_j}{d_j} - a_j\right)$$

$$= r\left[c_i\left(\frac{b_i}{d_i}\right) - a_i + c_j\left(\frac{b_i}{d_i}\right) - a_j\right] \qquad \text{(using 3-49)}$$

$$= r\left[\frac{b_i}{d_i}(c_i + c_j) - (a_i + a_j)\right]$$

$$= r\left[\frac{b_i + b_j}{d_i + d_j}(c_i + c_j) - (a_i + a_j)\right] \qquad \text{(using A-14)}$$

$$= r\,(b_i + b_j)\left(\frac{c_i + c_j}{d_i + d_j} - \frac{a_i + a_j}{b_i + b_j}\right)$$

Thus, criterion (b) is valid.

APPENDIX B. A LINEAR REGRESSION EXAMPLE

In this appendix we present an example of a simple linear regression problem which illustrates the care which must be taken in the use of regression.

Table B-1 presents boating accident statistics taken from CG-357 for the years 1965 through 1973. We determine a regression line for the variable y = annual fatalities, against the variable x = annual vessels involved in reported accidents.

Using standard regression equations (or, in this instance, a Texas Instruments SR-51A calculator), we determine the equation of this regression line to be:

$$y = ax + b = 0.142x + 680 \ . \qquad \text{(B-1)}$$

The standard deviations in x and y are:

$$s_x = 605.7 \text{ and } s_y = 147.4 \ .$$

The coefficient of correlation of x and y is given by:

$$r = \frac{as_x}{s_y} = 0.582 \ .$$

The unwary might proceed to use this regression line without further thought. However, we can test the significance of the correlation coefficient using the t-statistic:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \text{ , with n-2 degrees of freedom ,}$$

where n is the number of pairs of data points.*

We calculate:

$$t = 1.894 \ .$$

---

* See Reference 5 .

TABLE B-1. RECREATIONAL BOATING ACCIDENT STATISTICS

Year Reported	Vessels Involved in Reported Accidents	Reported Fatalities
1973	6738	1754
1972	5044	1437
1971	4915	1582
1970	4762	1418
1969	5239	1350
1968	5427	1342
1967	5274	1312
1966	5567	1318
1965	4778	1360

(From U.S. Coast Guard CG-357 statistical summaries.)

As we would expect a positive relationship between fatalities and vessels involved in accidents, we test the null hypothesis $r = 0$ against the one-sided alternative $r < 0$. A t-test indicates siginficance at just above the 5% level, so we should feel fairly safe in assuming there is a positive correlation, right? Wrong. Look at Figure B-1 in which the data points are plotted. If the 1973 data point is removed, an entirely different regression line is obtained. The equation of this line, based on 1965 through 1972 data is:

$$y = -0.170x + 2263 \ . \qquad \text{(B-2)}$$

The corresponding standard deviations and correlation coefficient are:

$$s_x = 298.7 \ , \quad s_y = 89.4 \quad \text{and} \quad r = -0.569 \ .$$

We now see the value of plotting the data points. The 1973 point is so atypical that it excessively influences the slope of the regression line.

This example should illustrate the importance of plotting data points whenever possible and of using tests to determine if the assumptions underlying the regression are satisfied. The reader is referred to References 6 and 7 for additional material on regression.

AD-A036 579 WYLE LABS HUNTSVILLE ALA F/G 12/1

REGULATORY EFFECTIVENESS METHODOLOGY. PHASE I RESEARCH.(U)

JUL 76 S B COHEN DOT-CG-42333-A

UNCLASSIFIED MSR-76-42 USCG-D-7-77 NL

2 OF 2

AD
A036579

END

DATE
FILMED
3-77

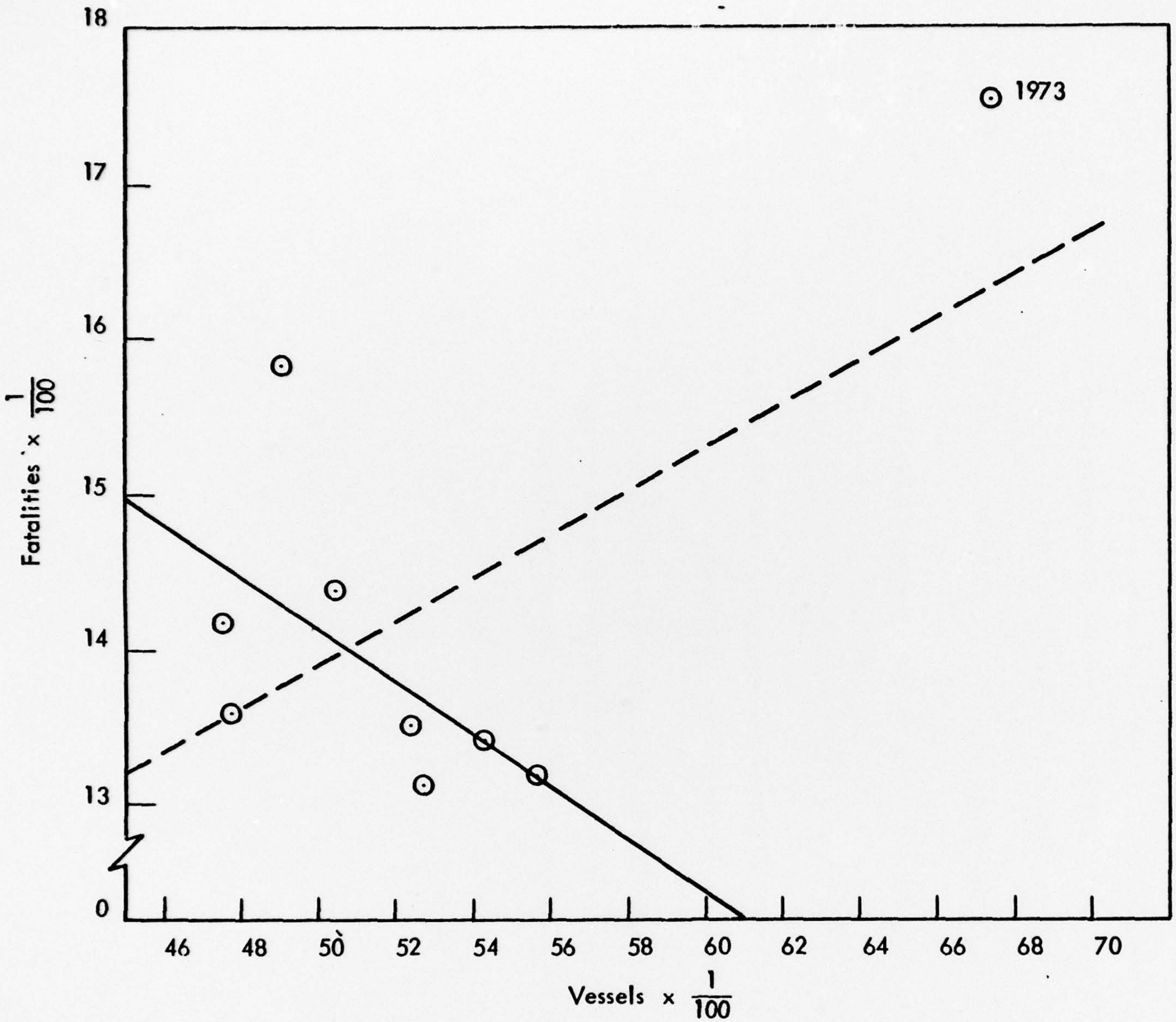


FIGURE B-1. FATALITIES vs VESSELS INVOLVED IN REPORTED ACCIDENTS 1965-1973

The solid line is the regression line calculated from all data points except 1973.
The dashed line is the regression line calculated from all points.